4-2016

Is metabolism goal-directed? Investigating the validity of modeling biological systems with cybernetic control via omic data

Frank T. DeVilbiss
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Frank DeVilbiss

Entitled

Is Metabolism Goal-Directed? Investigating the Validity of Modeling Biological Systems with Cybernetic Control via Omic Data

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Doraiswami Ramkrishna
Chair

John Morgan

Michael Gribskov

Joseph Pekny

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Doraiswami Ramkrishna

Approved by: John Morgan 4/5/2016

Head of the Departmental Graduate Program Date

IS METABOLISM GOAL-DIRECTED? INVESTIGATING THE VALIDITY OF MODELING BIOLOGICAL SYSTEMS WITH CYBERNETIC CONTROL VIA OMIC DATA

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Frank T. DeVilbiss

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2016

Purdue University

West Lafayette, Indiana

To Rongrong

ACKNOWLEDGMENTS

I could not have completed this dissertation on my own. There have been a great number of individuals who have helped me along the way on my academic journey and for whom I am very greatful.

I would like to first thank Prof. D. Ramkrishna for his invaluable guidance, mentorship and contribution of intellect. His openness to new ideas and well-developed intuition in all things modleing have allowed me to flourish in my investigations of nature. Along with this, his hands-off approach to advisement has given me a chance to explore without fear of constraint. A sincere thanks also goes out to Prof. J. Morgan who has been providing direction since my days as an undergraduate. His sharp mind has served as a critical reality check for my research from day one and I am happy that I was able to have my first research experience in his laboratory. Genuine appreciation also goes out to Professors Gribskov and Pekny who gladly accepted my request for them to serve on my committee. Both of you have helped me to become a better problem-solver.

Next, I would like to thank the many research group members who have given me direction. Thanks to Dr. Meenesh Singh and Dr. Hyun-Seob Song for taking me under your wings and directly mentoring me as an undergraduate. I am especially thankful for my experience with Dr. Song who has, to me, always been the model researcher - a paragon of hard-work, and dedication. Thank you also to my seniors Che-Chi Shu, Aravinda Mandli and Jayachandran Devaraj. My research group peers Conor Parks, Vu Tran, Parul Verma, and Akancha Pandey have been supportive peers throughout my tenure at Purdue.

Thank you also to the many with whom I've had the chance to collaborate. Dr. Mano Maurya, Dr. Shakti Gupta, and Prof. S. Subramaniam at UCSD have made

great contributions towards the modeling of arachidonic acid metabolism in this thesis. Their expertise has helped to make portions of this work possible. Prof. M. Raginsky at UIUC has been so helpful in the development of this model selection framework. Special gratitude also goes to Pablo Robles-Granda and Dr. Mohan Gopaladesikan who have donated much time to contribute to this project.

I have also been fortunate for the opportunity to collaborate with Mainak Chowdhury at Stanford on the use of data mining to improve patient outcomes. I am very thankful for our interaction and have learned much.

A heartfealt thank you goes to my family. To my parents, you taught me how to think like an engineer from day one and I have always been better for it. Thanks Ann for being the world's best sister and for the neverending moral support through these years.

Last, but certainly not least, I would not be the person I am today without the love and support of my dear Rongrong. Thank you for your many sacrifices to be at my side and for always pushing me to be a better person.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

DeVilbiss, Frank T. PhD, Purdue University, May 2016. Is Metabolism Goal-Directed? Investigating the Validity of Modeling Biological Systems with Cybernetic Control via Omic Data. Major Professor: Doraiswami Ramkrishna.

Cybernetic models are uniquely juxtaposed to other metabolic modeling frameworks in that they describe the time-dependent regulation of cellular reactions in terms of dynamic "metabolic goals." This approach contrasts starkly with purely mechanistic descriptions of metabolic regulation which seek to explain metabolic processes in high resolution - a clearly daunting undertaking. Over a span of three decades, cybernetic models have been used to predict metabolic phenomena ranging from resource consumption in mixed-substrate environments to intracellular reaction fluxes of intricate metabolic networks. While the cybernetic approach has been validated in its utility for the prediction of metabolic phenomena, its central feature, the goal-directed control strategy, has yet to be scrutinized through comparison with omic data. Ultimately, the aim of this work is to address the question "Is metabolism-goal directed? through the analysis of biological data. To do so, this work investigates the idea that metabolism is goal-directed from three distinct angles.

The first is to make a comparison of cybernetic models to other metabolic modeling frameworks. These mathematical formulations for intracellular chemical reaction networks range from purely mechanistic, kinetic models to linear programming approximations. Instead of comparing these frameworks directly on the basis of accuracy alone, a novel approach to systems biological model selection is developed. This approach compares models using information theoretic arguments. From this point of view, the model that compresses biological data best captures the most regularity in the data generated by a process. This framework is used to compare the flux

predictions of cybernetic, constraint-based and kinetic models in several case studies. Cybernetic models, in the test cases examined, provide the most compact description of metabolic fluxes. This method of analysis can be extended to any systems biological model selection problem for the purposes of optimization and control.

To further examine cybernetic control mechanisms, the second portion of this dissertation focuses on confronting cybernetic variable predictions with data that is representative of enzyme regulation. More specifically, the dynamic behavior of cybernetic variables, u_i, which are representative of enzyme synthesis control are matched with gene expression data that represents the control of enzyme synthesis in cells. This comparison is made for the model system of cybernetic modeling, diauxic growth, and for prostaglandin (PG) metabolism in mammalian cells. Via analysis of these systems, a correlation between the dynamic behavior of cybernetic control variables and the true mechanisms that guide cellular regulation is discovered. Additionally, this result demonstrates potential use of cybernetic variables for the prediction of relative changes in gene expression levels.

The last approach taken to test the veracity of cybernetic control is to develop a technique to mine objective functions from biological data. In this approach, returns on investment (ROIs) for various pathways are first established through simultaneous analysis of metabolite and gene expression data for a given metabolic system. Following this, the ROIs are used to determine a metabolic systems observed "goal signal." Gene expression data is then mined to select genes that show expression changes that are similar to the goal signal's behavior. This gene list is then analyzed to determine enriched biological pathways. In the final step, these pathways are then surveyed in the literature to establish feasible metabolic goals for the system of interest. This method is applied to analyze diauxic growth and prostaglandin systems and generates objective functions that are relevant to known properties of these metabolic networks from the literature. An enhanced understanding of metabolic goals in mammalian systems generated by this work reveals the potential utility of cybernetic modeling in new directions related to translational research.

Overall, this investigation yields support of the notion of dynamic metabolic goals in cells through comparison of metabolic modeling approaches and through the analysis of omic data. From these results, a lucid argument is made for the use of goal-directed modeling approaches and a deeper understanding of the optimal nature of metabolic regulation is gained.

1. INTRODUCTION

Cybernetic models have been used, for a period that now spans over three decades, to predict a variety of metabolic phenomena. From their initial starting point in describing diauxic growth in a highly lumped fashion, they have been modified and advanced to provide descriptions of complex metabolic behavior including intracellular reaction fluxes and support efforts in metabolic engineering.

1.1 "Are cybernetic control variables a good approximation of cellular regulation?"

The distinguishing feature of cybernetic models of metabolism is the inclusion of control variables that represent the modulation of cellular chemical reactions towards goals. While cybernetic models have had much success in predicting a variety of metabolic phenomena, this control policy merits further scrutiny and validation. This dissertation seeks to test the veracity of cybernetic control as a means of describing the bulk of regulatory actions undertaken by cells. To examine cybernetic control further, several approaches will be taken.

The first is to make a comparison of cybernetic models to other metabolic modeling frameworks. The mathematical formulations of these descriptions of intracellular chemical reactions range from purely mechanistic, kinetic models to linear programming ones. Instead of comparing these frameworks directly on the basis of accuracy alone, a novel approach to systems biological modeling is developed. This approach compares models using information theoretic arguments. Through the lens of this approach, the model that compresses biological data best, captures the most regularity in the data that the process generates. In establishing the fact that cybernetic models provide the most compact description of metabolic dynamics relative to other models, the control goals that make them unique are validated in their pragmatism. They

provide accurate predictions that require less effort for parameterization and model specification. This method of analysis can be extended to any systems biological model selection problem for the purposes of optimization and control.

To further verify cybernetic control, the second portion of this dissertation focuses on the comparison of cybernetic variables with data that is representative of enzyme regulation. More specifically, the dynamic behavior of cybernetic variables, u_i, which are representative of enzyme synthesis control are matched with gene expression data that represents the control of enzyme synthesis in cells. This comparison is made for the model system of cybernetic modeling, diauxic growth, and for prostaglandin (PG) metabolism in mammalian cells. From this analysis, it becomes clear that there is a strong correlation in the dynamic behavior of cybernetic control variables and the true mechanisms that guide cellular regulation. This makes a strong argument for the validity of cybernetic control mechanisms. Additionally, this analysis demonstrates how cybernetic variables could be used to make predictions about the relative changes in gene expression data for a given application.

The last approach taken to test the veracity of cybernetic control is to make an attempt at mining objective functions from biological data. If one can determine feasible metabolic goals through unbiased analysis of cellular observations, a strong argument for goal-directed metabolism is made. In this approach, returns on investment (ROIs) for various pathways are first established through simultaneous analysis of metabolite and gene expression data for a given metabolic system. Following this, the ROIs are used to establish a metabolic systems observed "goal signal." Gene expression data is then mined to determine genes that show expression changes that are similar to the goal signal's behavior. This gene list is then analyzed to determine enriched biological pathways. These pathways are then surveyed in the literature to establish feasible metabolic goals for the system of interest. This method is applied to diauxic growth and prostaglandin systems and is used to generate verifiable objective functions.

Overall this dissertation demonstrates the veracity of cybernetic control through comparison of metabolic models and also through the analysis of omic data. In accomplishing this task, a lucid argument is made for the use of goal-directed modeling approaches. Moreover, a deeper understanding of the optimal nature of metabolic regulation results.

1.2 Organization of Dissertation

The organization of this dissertation is as follows. In chapter 2, extensive coverage of literature relevant to metabolic modeling is made. This should serve to provide context to the cybernetic approach and the research questions at hand. In chapter 3, an information theoretic model selection framework is developed for systems biological applications. This helps to show that cybernetic models provide a more succinct description of metabolic dynamics than other approaches. Chapters 4 and 5 cover the validity of the matching cybernetic variable via its comparison with time-dependent changes in gene expression data. Chapter 6 establishes a case for mining omic data for objective functions and for improved parameterization approaches for cybernetic models through a discourse on some shortcomings in the framework with making postulates about objective functions. Following this, Chapter 7 introduces a method to mine omic data for metabolic goals. Chapter 8 provides for an improved method of parameterizing cybernetic models that is computationally efficient. Chapter 9 summarizes the findings of this work and Chapter 10 covers some recommendations for future work.

2. BACKGROUND AND LITERATURE REVIEW

Many formulations have been developed to model and characterize metabolic processes. Among these, the cybernetic framework stands out in its ability to describe the evolution of metabolism using the concept that the cell has some preprogrammed artificial intelligence coordinating the complex mechanisms that drive metabolic changes. This behavior comes about through the iterative refinement of organisms via an evolutionary process which converges upon this seemingly optimal behavior. From the Darwinian viewpoint, organisms that are more refined towards achieving goals related to their survival are better equipped to pass on their genes to the next generation. Behaviors that demonstrate optimality, in other words, those which conform to perfected survival strategies, will be the most competitive and outlasting. The following review of relevant research seeks to provide the reader with an understanding of prior work to date in the field of cybernetic modeling and to give context to the cybernetic framework by discussing other metabolic model formulations.

2.1 Cybernetic Modeling

Cybernetic modeling of metabolism, at its core, embodies a framework of ordinary differential equations that describe the time-dependent evolution of metabolite and enzyme concentrations. In cells, these changes in concentrations, both inside and outside of the cell, are governed by the directed actions of a host of complex biological machinery. The main feature of cybernetic models is the inclusion of control variables that interpret the sum of these regulatory processes as actions that seek a goal related to the cell's survival. Without knowing the full detail of all mechanisms related to biological control, these variables endow the cybernetic approach with the ability to describe changes in metabolites in a host of biological systems with accu-

racy. Monod's seminal work upon the diauxie provides a cornerstone for the study of biological regulation and inspired the first cybernetic models [1]. Monod subjected *Bacillus coli* cells to environments with multiple carbon substrates upon which the cells could grow. From this, the changes in substrate and cell concentrations were tracked over time and it became apparent that the cells preferred some substrates over others and would not simultaneously consume carbon sources simply according to the relative concentrations of substrates available. Describing the diauxie was the first system described by the cybernetic approach as it embodies a foundational study of regulatory phenomena in biology.

2.1.1 Lumped Cybernetic Models

Cybernetic modeling started almost 30 years ago with the goal of modeling microbial growth on multiple substrates using the control variable u_i to describe resource allocation for enzyme generation for each substrate's digestion [2]. This simple model was then expanded upon in Kompala's work which included the variable v_i as an approximation of the modification of enzyme activity for the digestion of different substrates [3]. Following this, constitutive enzyme synthesis was incorporated into the framework [4]. In other work by the same authors, provisions were added to describe maintenance, non-growth associated processes, within the cell [5]. This model was able to explain the behavior of cell cultures at low growth rates on single and multiple substrates. However, it was incapable of predicting transients in continuous culture. To remedy this, Baloo and Ramkrishna included specifications for cellular resources that become limited during the lag period in step-up experiments, namely transcriptional resources for enzyme production [6, 7]. Also included were control variables that described competition among growth and maintenance processes on single and multiple substrates.

Pushing the framework forward, using dynamic analysis, Narang observed that Kompala's model was unable to describe the simultaneous uptake of substrates and organic acids given their structural rigidity [8]. This was in part due to the assump-

tion that growth rate on a mixture is never greater than the maximum growth rate of individual substrates. Moreover, the structure of Kompala's model regarded the formation of biomass from substrate as a single step. Reworking this framework to include generic M_1 and M_2 biomass precursors, Ramakrishna et al. were able to describe simultaneous uptake patterns in mixtures of substrates [9]. Until Ramakrishna's model, the cybernetic approach had been highly lumped. The inclusion of biomass precursors marks a departure in cybernetic approach from simple substrate pathways towards applications in metabolic engineering and systems biology because it describes the formation of biomass in multiple steps.

2.1.2 Structured Cybernetic Models

The subsequent class of cybernetic models decomposed the aforementioned lumped pathways into functional units that reference individual parts of the metabolic networks. The first effort in this direction sought to apply cybernetic principles related to resource competition among generic features present in metabolic networks including linear, branched and cyclic pathways [10]. In principle, these features could be modeled with local objective functions and then integrated together to develop models for larger networks.

Expanding the scope of this work with the explicit desire to apply the cybernetic approach to metabolic engineering applications, Varner and Ramkrishna postulate that individual enzymes are controlled by the product of both local and global objective functions where the local objectives determine resource allocation within pathways while global objectives determine regulation between different pathways [11,12]. These local and global control policies complement one another towards the cell's ultimate goal of survival. Using these ideas, a framework for predicting the behaviors of genetically modified microorganisms is proposed using metabolic control analysis to analyze the sensitivities of enzymes. This framework was used to describe modification of amino acid synthesis within *Cornyebacerium lactoferum* using overexpression [13]. In a later work, a model was developed to describe gene knockouts

in *Escherichia coli* which provided an accurate prediction of almost all flux ratios in central carbon metabolism's split points. Following this, Namjoshi et al. showed how these structured cybernetic models can be used to predict steady-state multiplicity for hybridoma cells in chemostat reactors using a highly simplified reaction network [14, 15].

Within Young et al.'s work, a refined the notion of global control variables was introduced to depict the regulation of various modules of metabolism in the central carbon metabolism of *E. coli* [16]. Essentially, the global control variables coordinate the control of metabolism to maximize the rate of formation of various growth precursors which results in the dynamic maximization of growth. This model is then able to predict the behavior of wild-type and acetate knockout strains of bacteria. The key element of this work is how it divides the metabolic network into related components which are similarly controlled.

Also of note, Young et al. analyzed the cybernetic framework underneath the microscope of optimal control theory and compared a variety of control policies in order to demonstrate their relative merits related to growth in mixed substrate cultures [17]. Of these policies, the enzyme weighted matching and proportional laws proved the best fit to data demonstrating the validity of cybernetic models.

2.1.3 Pathway Cybernetic Modeling

While structured cybernetic models could be used to model problems relevant to metabolic engineering, their formulation was cumbersome in that the metabolic network analysis underpinning their formulation, first prescribed by Straight et al. [10], relied upon the modeler's intuition when dividing the network into various modules. This subjective approach cannot easily be extended to an arbitrarily defined network which is a significant limitation. To more systematically incorporate reaction networks into cybernetic regulation, a new class of cybernetic models were developed using elementary flux modes (EFMs) by Young et al. [16]. EFMs describe the fundamental, non-decomposible pathways through a metabolic network from beginning

point to end point [18]. Given that the generation of EFMs entails an unbiased and comprehensive breakdown of the network into its basic functional units, modeling networks on the basis of EFMs provides a more consistent and generalizable approach. In the first cybernetic model using EFMs, Young et al. decomposed a simplified network of 12 reactions into 8 EFMs and described the regulation of these pathways in order to achieve the objective of maximizing biomass [16]. Note that in order to apply this method, one must use a lumped biomass reaction which represents the conversion of precursors, ATP and NADH into a unit of biomass.

Expanding upon this framework, Kim et. al described a similar model termed Hybrid Cybernetic Model (HCM) in which the maximization of substrate uptake was the objective function [19]. The main difference between HCM and Young's model is that the former makes use of a quasi-steady state hypothesis about the concentration of metabolic intermediates which relieves the HCM framework of its need for kinetic parameters related to intermediate reactions. To implement the objective function of substrate uptake maximization, in terms of the cybernetic variable, the coefficient for substrate uptake in each EFM was used to weight the return on investment. This approach yielded a slightly better fit to the data as compared to Young's model for the simplified 12 reaction network. Also, the expression of each EFM in terms of the cybernetic variable profile was also different in Kim's model. The HCM framework was then applied to a more complicated metabolic network for *E. coli* strain GJT001 which included 23 different reactions. This more complicated network incorporates the pentose phosphate pathway (PPP) and oxaloacetate formation as an intermediate in the production of succinate. The lumped biomass reaction contains 14 biosynthetic reactants compared to the 3 reactants in the simpler network. This more detailed model helps to demonstrate HCM's potential to be applied to more complex reaction networks.

HCM has also been used to identify a multiplicity of steady states in a chemostat reactor for the GJT001 strain growing on mixtures of glucose and pyruvate [20]. The model was composed of EFMs that described the sole uptake of glucose, pyruvate

as well as mixed modes for the combined uptake of both substrates. Depending on which direction the steady state was established from (i.e. from low dilution rate to high or the opposite), different combinations of EFMs would be activated. Mathematically, this is the result of the nonlinear nature of the equations describing the biological phenomena. This steady state multiplicity was later validated via experimental observation in [21].

One of the main drawbacks of HCM is the number of parameters required to describe the unregulated kinetics of each pathway. As the number of reactions in the network increases the number of EFMs undergoes a combinatorial explosion. In order to address this, Song and Ramkrishna proposed Yield Analysis (YA) to pick out the EFMs that defined extreme edges circumscribing the convex hull in yield space [22]. Using YA, one could drastically reduce the parameterization of HCM by only considering these edges as many EFMs can be represented as a convex combination of the selected subset of EFMs and still retain an ability to model the majority of the possible metabolic phenotypes using all EFMs.

Further expanding upon this idea that the pathway-based cybernetic model can be parametrically simplified, Song and Ramkrishna introduced a Lumped Hybrid Cybernetic Model (LHCM) in which EFMs for specific substrates are weighted and summed into a single flux mode based upon each EFM's ability to generate ATP for maintenance and biomass for growth [23]. Using this lumping scheme, the total set of EFMs is first divided into substrate specific groups. Within these groups, a structural return on investment (s-ROI) is calculated based off of the yield of biomass and ATP in each EFM. The EFMs within the substrate group are then combined with a weighted sum based off of each EFM's s-ROI. This drastically reduces the parameterization of this type of model however there is a tradeoff in that the modes within each substrate group are not subject to dynamic control only the lumped EFMs have cybernetic variables regulating their changes in expression and activity over time. Another issue with the original formulation of Song and Ramkrishna was the fact that the lumping process severely reduces the amount of possible metabolic

states that can be modeled. To address this, Song and Ramkrishna added tuning parameters in the lumping process that are designed to fit the lumped elementary modes to yield data [24].

L-HCM has been used in a variety of applications relevant to metabolic engineering. It has been used to model the behavior of different strains of *E. coli* with the same general metabolic network. It has also been used to model the behavior of mutant *E. coli* strains using limited wild-type data demonstrating its relevance to metabolic engineering [25]. L-HCM has also been used to calculate the metabolic fluxes of reactions related to the complex growth phases of Shewanella oneidensis (MR-1) on lactate [26]. When compared to flux predictions generated by Flux Balance Analysis (FBA), the L-HCM model demonstrated better predictive ability. These practical applications of the L-HCM framework help to display the validity of lumping as a tool to simplify the parameterization of pathway based cybernetic models.

2.2 Other Metabolic Modeling Tools

There are a wide variety of other modeling tools available to one who would like to describe various aspects of metabolism. Kinetic models, describing the rates of different metabolic reactions as a function of metabolite and enzyme concentrations, are often used to describe the dynamics of metabolism. Other approaches that mainly consider stoichiometric information, termed constraint-based models (CBM) like flux balance analysis (FBA) offer one an ability to quickly ascertain general features related to metabolic fluxes through a given network. Depending on the system, certain models are more appropriate than others. The discussion that follows intends to outline the general features of analytical tools for metabolic engineering that intersect in functionality with the cybernetic framework.

2.2.1 Kinetic Models

Kinetic models are formulated using a system of first order ODEs that track the changes in concentration of various chemical species. The rates of chemical reactions are determined by appropriately formulated rate expressions that dictate the speed at which chemical species are transformed from one type to another. Typically, rate expressions are formulated in a manner that captures the behavior of enzymes. Approximations of enzymatic behavior can range from simple linear kinetics or Michaelis-Menten kinetics to complex forms like Monod-Changeux-Wyman kinetics [27]. While linear kinetic models may be considered to be an oversimplification of enzymatic reactions, they lend themselves to convenient analytical techniques that determine criteria such as stability around steady states [28]. Nonlinear kinetic descriptions of enzymes, such as Michaelis-Menten kinetics, also approach linear behavior when the substrate concentration is significantly lower than the reaction's Michaelis-Menten constant.

As time moves forward, progressive degrees of complexity are being incorporated into kinetic models of biological phenomena. Provisions for regulation mechanisms in metabolism like allosteric interactions, transcriptional regulation and signaling serve to shed light on the feedback structures within metabolic processes [29, 30]. While there is an increasing breadth of omic data available to aid in the development of kinetic models, the determination of kinetic parameters is still a pressing problem [31]. Limitations in the amount and resolution of data make nonlinear regression difficult. Also challenging is the fact that modeling intuition is needed when determining which interactions are necessary in a model and which ones can be lumped together. Ensemble modeling approaches can help the modeler determine which interactions are necessary to include but require significant computational input to implement [32]. Also adding to the difficulty of kinetic modeling is the fact that, *in vivo*, many parameters related to biological phenomena change from what is measured *in vitro*.

2.2.2 Constraint-Based Models

CBMs on the other hand, offer quick solutions to determine metabolic fluxes with minimal parameterization. Using the stoichiometric matrix describing metabolic reactions, CBMs can be used to analyze theoretical yields of various pathways and determine metabolic overflows. FBA is formulated as a linear programming problem designed to find the metabolic pathway that optimizes some objective function. These objective functions can take the form of maximization of biomass yield in single cellular organisms [33] or minimization of reactive oxygen species (ROS) in hybridoma cells [34]. To determine the behavior of a modified network, a relevant constraint-based approach is minimization of metabolic adjustment (MOMA) which uses quadratic programming to solve an objective function that seeks a flux distribution that minimizes the norm between an original flux distribution and a perturbed one [35]. This approach is useful for predicting the phenotypes of various knockout strains. Another useful feature of MOMA is that it does not require the input of substrate uptake rates to make predictions. Similar to MOMA, regulatory on/off minimization predicts alternate flux distributions for gene knockouts. Instead of minimizing the total change in all fluxes, ROOM identifies a solution of the lowest number of individual fluxes changed using MILP in which a vector of Boolean variables classifies whether the vector of fluxes has been changed from its original formulation and is summed [36].

Investigating the validity of different objective functions to be used in the FBA approach, specifically maximization of growth rate, Schuetz et al. generated a Pareto-optimal surface to model the tradeoffs between a set of objective functions [37]. Comparing CBM predictions determined by different combinations of the objective functions of growth maximization, ATP maximization and minimizing the sum of fluxes with actual data from nine different organisms, it is shown that metabolic data is close to the Pareto-optimal surface determine by combinations of objective functions.

CBMs are also readily applied to large, genome-scale networks for quick analysis of metabolic states.

The main weakness of CBMs is the exclusion of the dynamics and regulation inherent to biological systems [38]. Attempting to add time-dependent considerations into this framework, dynamic Flux Balance Analysis (dFBA) adds either a static or dynamic optimization approach into calculating the flux distribution [39]. However, constraining these predictions is the fact that the objective functions are calculated towards the maximization of yield given a network structure and do not make use of actual kinetic information regarding the network. Another weakness, as compared to kinetic models, is that CBMs do not accomodate intracellular metabolic concentrations.

2.2.3 Biochemical Systems Theory

Biochemical systems theory (BST) is a simple modeling approach that is able to capture the full dynamics of complicated systems [40]. The change in each metabolite is modeled using easily analyzed power-law expressions that describe the rates of formation and degradation of different species in the system. All changes in metabolites $\dot{x}_i$ are described with expressions of the form:

$$\dot{x}_i = \alpha_i \prod_j x_j^{g_{ij}} - \beta_i \prod_j x_j^{h_{ij}}$$

The parameters α_i and β_i are constants of formation and degradation and are constrained at values above zero. The exponents g_{ij} and h_{ij} in the model describe x_j's participation in the formation or degradation of species x_i. Even signaling effects like inhibition or activation of enzymes can be modeled for species x_j by specifying an appropriate g_{ij} or h_{ij}. Using BST, one can model the response of the system to perturbations, steady states, and complex oscillatory phenomena. While easily generated, a major drawback of BST is that it is unclear if the model can be used over a wide range of physiological conditions, especially those far from what was

used to parameterize the model. Moreover, the physical meaning of the exponents is unclear. BST has been used for a variety of applications including the modeling of the growth of *E. coli* on glucose [41].

2.2.4 Metabolic Control Analysis

Toward the metabolic engineering of different biochemical pathways within cells, there exists a strong desire to effectively manipulate elements of the pathways for certain applications. While one may wish to manipulate a biochemical pathway to increase or decrease its throughput, manipulation of the rate-limiting step is not necessarily the best way to do this. Unforeseen control interactions or pathway dependencies can prevent one from truly achieving this goal. Metabolic control analysis (MCA) is designed to systematically and qualitatively gauge the amount of control each enzyme and effector has on the overall flux of a pathway [42, 43]. MCA is useful in that it can help one determine the most appropriate modification for a pathway, however the predictions it provides are limited to conditions from which the control coefficients are calculated. These control coefficients can also be generated from a kinetic model describing the dynamics of a pathway.

3. MODEL SELECTION PROBLEM: RELATIVE COMPRESSION OF SYSTEMS BIOLOGICAL DATA BY CYBERNETIC MODELS

Chapter Summary

The field of systems biology thrives upon the use of models to organize biological knowledge and make predictions of complex processes that are hard to measure. When attempting to generate model descriptions for metabolic systems, one arrives at a crossroads. A variety of mathematical explanations are available for metabolic data with varying degrees of resolution from simple to complex. Biological modelers often rely upon subjective arguments to choose one framework over another. While there is no universal rule to determine the absolute utility of a model, certain metrics founded on information theoretical principles, demonstrate promise in providing a coherent, rational, and objective basis for addressing this model selection problem in systems biology. A model seeks to capture the regularity in biological data. Models that best capture regularity in data without excessive complexity are the most useful for applications in optimization and control. To demonstrate the efficacy of such an approach, several metabolic model selection scenarios are investigated. This work develops the argument that information theoretic model selection metrics should be extended to non-nested model comparison applications in systems biology. It also makes a novel comparison of kinetic, constraint-based and cybernetic models of metabolism based not only on model accuracy, but also model complexity. The results show the strengths of Lumped Hybrid Cybernetic Model (L-HCM) and Flux Balance Analysis (FBA) for applications in steady state flux prediction. Also, the Hybrid Cybernetic Model's (HCM) merit in the modeling of dynamic changes in fluxes is also established.

3.1 Introduction

When given an arbitrary set of data, one can generate a host of different mathematical descriptions for it. Metabolic systems are no exception and embody an important branch of systems biological study. In order to predict the effects of perturbations to metabolic networks such as deleting genes or inhibiting enzymes, it is useful to first use a model to understand, without additional experimentation, the effects of such modifications. To model the changes in metabolic systems, one can select kinetic, constraint-based or cybernetic formulations. Each of these metabolic models are unique in formulation and are widely used for similar goals.

In very general terms, the utility of a model is derived from its ability to describe regularity in data. Regularity, or coherence in a set of data, means that the data is generated as the result of some intelligible process [44]. For metabolic flux data, each type of model is able to capture the coherence of metabolic processes to a certain degree. These models also have varying degrees of complexity which are used to explain the behavior of data. To establish which model is best for the purposes of optimization and control, it is proposed that models are selected on the basis of how well they are able to capture the regularity of data without being excessively complex.

In this work, a number of widely used mathematical approximations of metabolic systems are compared according to the ability of each to capture regularity in data. While there is much discourse on the merits of each modeling framework [45–49], no systematic method has been implemented to quantitatively and simultaneously measure the relative accuracies and complexities of metabolic models. This work, applies information theoretic metrics, well-known in other fields, to address this problem. Treating each model as an entity that compresses data for communication through a channel, one can quantitatively evaluate how well a model balances accuracy and complexity. Models that accurately reproduce data with low complexity require less information to communicate and embody a more compressed description of a pro-

cess's data. This method of evaluation is especially useful in situations where model formulations are vastly different (i.e. non-nested).

The establishment of the model that best minimizes penalties for both error and parameterization is deemed to be the best model for an application related to the optimization and control of biological systems. Restated, there is a point in diminishing returns for model complexity. Additional parameters may enhance accuracy, but each additional parameter has an intrinsic cost associated with it.

A set of 4 distinct models of metabolic fluxes are judged in their ability to describe metabolic reaction rates at a given steady state. Following this, dynamic metabolic models are compared in their ability to predict changing metabolic fluxes. These dynamic models of metabolic fluxes have never been compared in this objective fashion. Neither have such a wide range of steady state descriptions of metabolic fluxes been compared. The application of these metrics in these scenarios helps to establish a new way of thinking about metabolic model selection. Moreover, a quantitative framework for comparing non-nested biological models is necessary to introduce to the field of systems biology.

3.2 Methods

3.2.1 Theory

To develop the model selection framework, it is first useful to review some basic tenets of information theory for those who might be unfamiliar. In the field of communication, signals that are being passed through a channel are analyzed and compressed depending on how much regularity is present in a given message. Compression is a useful tool in that shorter messages lead to faster communication. For example, consider the two messages below:

a. 011010111010010001011

b. 000010000000000000000

Sequence (a) is generated from some arbitrary process where either 0 or 1 share equal probabilities of occurence. Sequence (b) on the other hand is generated from a process where the probability of 1 is 1/20. To communicate either sequence, one could send each 1 or 0 value individually or one could compress the information down into a shorter sequence. Given the fact that either ones or zeros are equally probable in (a), it is virtually incompressible. Sequence (b), on the other hand, can be shortened in a number of ways.

For example, sequence (b) can be simply described by specifying the position of the single one rather than the whole sequence assuming the decoder understands the compression scheme. In case (b), one could rewrite the sequence as "5" specifying the location of the single 1. For this 20 bit sequence, specifying the position of a single one in any of 20 possible locations requires up to 5 bits ($\lceil(log_2(20)\rceil$) instead of 20 for compressing any position in the sequence of 20 bits. This represents a compression factor of 0.25 compared to communicating the entirety of the original sequence. Other, more efficient coding schemes are possible and the fundamental limit of compression of these data sequences is quantified using Shannon's entropy [50]

$$H(x) = \sum_i p_i \log(p_i), \tag{3.1}$$

in which p_i represents the probability of a 0 or 1 occurring in the sequence. The motivation for compression is increasing the overall rate of communication. The more compressed a message is the less time is spent communicating it. In terms of entropy, the highest entropy sequence will consist of bits generated by the method of (a). In the same way that a message can be compressed by a proper coding scheme, we can say that biological data can be compressed by a model. It is here where the minimum description length princple (MDL) becomes useful in that one can reinterpret the model selection problem as one of data compression [44]. The aim is to shrink the data D into some D' from which D can be perfectly reconstructed after compression. For some model M, there is a length of the data $L(D')$ that is determined as

$$L(D') = L(D|M) + L(M), \tag{3.2}$$

where $L(D|M)$ expresses the data in terms of the model and $L(M)$ is a description of the model's complexity [44, 51]. The term, $L(D|M)$, accounts for the extra information that needs to be transmitted in order to describe the model prediction's distance from the real data. For example, if the model prediction comes close to the data, less information is needed to communicate the model error than if the prediction is far from the data in the same way that smaller integers can be communicated by fewer bits than larger ones (e.g. in the case of integers, the number 2 can be encoded into binary as "10" at 2 bits vs. 20 as "10100" at 5 bits). The complexity term $L(M)$ defines the amount of information that must be communicated in order to describe the model and is typically defined by the number of parameters in the model. Together, the model and a specification of its error can be used to perfectly reconstruct the data from D' to D.

To represent $L(D')$ for some model, one can apply metrics like Akaike's Information Criteron (AIC) [52] or Schwarz's Bayesian Information Criterion (BIC) [53]. These information theoretic metrics take the form of either

$$AIC = n\log(\hat{\sigma}^2) + 2k, \tag{3.3}$$

or

$$BIC = n\log(\hat{\sigma}^2) + k\log(n). \tag{3.4}$$

In the above, $\hat{\sigma}^2$ is the error of the model, k is the number of parameters within the model and n is the number of data points that the model approximates. Note that these metrics are valid asymptotically and correction factors are applied in the case of limited data. These correction factors increase the penalty on extra parameters when there are fewer data points.

AIC is formulated using Kullback-Liebler divergence and seeks to select a candidate model that best describes reality. This is due to the fact that KL divergence is a measure of the extra information needed to transmit some information using a model distribution as compared to some real generating distribution [54]. BIC, built from a Bayesian arguments, seeks to select a "true" model from a possible set of mod-

els. More specifically, it applies a likelihood function to gauge the probability that a model is true given some observed data. AIC penalizes parameters less severely than BIC which means that BIC tends to favor simpler models than AIC. For derivations of these metrics, one should consult [52, 53].

Both metrics above are made from different arguments, but they embody a similar principle. A model that optimally describes the data will strike a balance between accuracy and complexity.

To further explain these metrics in the context of a communication problem, consider the transmission of the model in the place of the raw data through a communication channel. In order to communicate the model itself, the parameters need to be transmitted with a certain accuracy. Encoding parameters to a precision of $\delta_m = 1/\sqrt{n}$ is the most reasonable way to do this as $1/\sqrt{n}$ represents the magnitude of estimation error on the parameters themselves [55]. To transmit a model's k parameters to this precision through the channel, one will need to use $-k \log_2 1/\sqrt{n}$ bits which makes up the latter portion of BIC.

The error of the model will also require communication which is approximated by the mean squared error for the data set. Mean squared error represents the average magnitude of error in the description of each data point. For a particular data point, magnitude of error is relevant because, as stated previously, larger numbers require more bits to communicate. When considering these model metrics in the context of data transfer, they are formally referred to as two-stage description length or two-stage MDL. Note that two-stage MDL has the same form as BIC.

To further illustrate the use of these information criteria, let us consider some arbitrary data set as shown in figure 1. There are a range of polynomial models that one could use to fit the n data points from an overly simple linear model of order 1 in (a) to an over-fitting Lagrange polynomial of order $n-1$ in (c). While the Lagrange polynomial captures the data with no error, the third order polynomial has a better qualitative fit to the data.

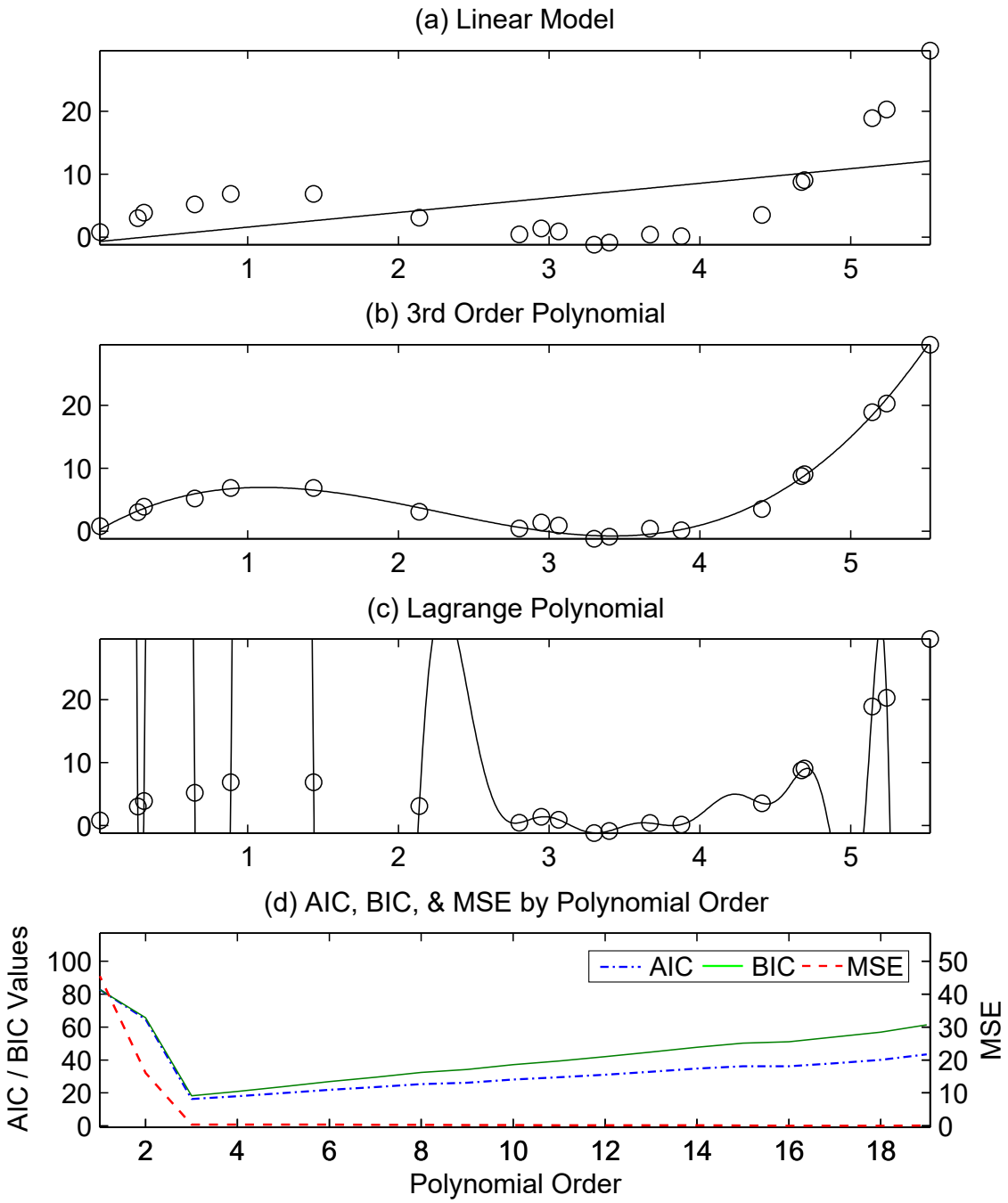

Figure 3.1.: Example of model selection problem with polynomials. (a) is the fit of a linear polynomial for the data. (b) is the fit of the third order polynomial. (c) is the fit of a Lagrange polynomial of order $n-1$. (d) shows the behavior of MSE in red while AIC and BIC are shown in blue and green respectively for each order of polynomial. Note that both metrics are minimized for the 3rd order polynomial model.

When using a model for the purposes of optimization and control, one desires a model that will be accurate without being overly complex. By applying AIC and BIC to the analysis of biological models, one can gain a relative sense of a model's balance between accuracy and complexity relative to other descriptions of the data. An important advantage offered by these metrics is that they can be used to compare non-nested models. In other words, these information criteria are valid for comparing models of vastly different formulations such as constraint-based and differential equation models.

The model selection approach highlighted contrasts with other methods of model selection that focus merely on accuracy. Approaches such as cross-validation are useful for comparing models but do not offer any insight into the relative complexity of different models. Other methods that seek to prevent overfitting such as regularization are not necessarily useful for the applications discussed here.

3.2.2 Metabolic Models

To demonstrate the value of these information criteria for biological modeling applications, a set of models will be developed to describe the same set of biological data. This set of models includes kinetic, constraint-based and cybernetic models. These models can all be used to predict metabolic fluxes, or the rates of different intracellular chemical reactions, in cultures of cells growing on different carbon sources. These models are all developed using experimental data representing quantities such as the dynamic changes in concentration of carbon sources, and the growth rates of cells to arrive at these predictions. Each model has varying complexity and predicts metabolic fluxes with different amounts of error. It is the goal of this work to show how well each model balances between these.

To understand the general features of each modeling framework, the structure of all three classes of models will be highlighted. To start, one must consider the nature of metabolic systems. They are composed of connected chemical reactions that form

networks. This network is articulated in a stoichiometric matrix $\mathbf{S}$ of m metabolites by n reactions. To model the changes in extracellular species, one can use

$$\mathbf{x} = \begin{bmatrix} \mathbf{s} \\ \mathbf{p} \\ c \end{bmatrix} \tag{3.5}$$

where $\mathbf{s}$, and $\mathbf{p}$ are vectors of n_s substrates, n_p products respectively, and c is the concentration of cells often referred to as biomass. Combining the extracellular variable $\mathbf{x}$ with a vector for the cell density normalized intracellular components $\mathbf{m}$ yields an expression that describes the time rate of changes of extracellular and intracellular variables

$$\begin{bmatrix} \frac{1}{c}\frac{d\mathbf{x}}{dt} \\ \frac{d\mathbf{m}}{dt} \end{bmatrix} = \mathbf{Sr}. \tag{3.6}$$

Above, $\mathbf{r}$ represents the rates of metabolic reactions or fluxes. In the kinetic model, this differential expression is solved using expressions for $\mathbf{r}$ that approximate the fluxes of the chemical reactions as a function of $\mathbf{x}$ and $\mathbf{m}$. These flux expressions typically use Michaelis-Menten kinetics such as

$$r_i = \frac{V_i^{max} m_i}{K_i + m_i}, \tag{3.7}$$

where V_i^{max} and K_i are the maximum reaction rate and saturation constants respectively. These parameters rely on experimental data and can change significantly for different reactions. Given that metabolic networks can be composed of thousands of reactions, kinetic models can be quite complex. Also, the kinetics used can also include enzyme influences and reaction inhibition. Kinetic models are typically very high resolution pictures of cellular processes.

Constraint-based models such as Flux Balance Analysis (FBA) embody a much simpler approach to predicting metabolic fluxes [33]. FBA makes two major assumptions to do so. One is that intracellular metabolites are at some pseudo-steady state or

$$\frac{d\mathbf{m}}{dt} = 0 \tag{3.8}$$

The other is that the cells organize their metabolic fluxes to optimize some objective function. This objective function typically takes the form of maximizing the yield of biomass. This objective function will be used in all of the proceeding scenarios. FBA is written as an optimization problem as

$$\begin{aligned} \max \quad & J = \mathbf{c^T r} \\ \text{subject to} \quad & \mathbf{Sr} = \mathbf{0} \\ & \mathbf{a} < \mathbf{r} < \mathbf{b} \end{aligned} \tag{3.9}$$

and can be solved using LP. Above, the product $\mathbf{c^T r}$ represents the combination of fluxes that are maximized. Fluxes are constrained to be in the null space of $\mathbf{S}$ and must satisfy a specified set of upper and lower bounds refered to as $\mathbf{b}$ and $\mathbf{a}$. These lower and upper bounds are determined by experimental evidence as well as thermodynamic constraints on the reactions (i.e. some reactions only work in the forward direction). The experimental evidence is typically used to constrain the uptake rates of substrates consumption and product formation in the model. Other intracellular constraints can be used, but these quantities can be difficult to measure. To model the dynamic changes of fluxes in this work, the static optimization approach will be used from dFBA where a model will be used to approximate the changes in constraints over time [39, 56].

Cybernetic models use dynamic objective functions that optimize the system to achieve goals at each time through the inclusion of control variables that regulate enzyme synthesis and activity. Instead of exhaustively describing the kinetics of each reaction as the kinetic model does, hybrid cybernetic models (HCMs) decompose the reaction network into a set of pathways or macroscopic reactions termed elementary modes (EMs) [19] that are expressed at varying levels over time. To do so, the pseudo-steady state assumption must be made like in FBA. The flux vector can be decomposed into a set of rates through the EMs as

$$\mathbf{r} = \mathbf{Z^T r_M} \tag{3.10}$$

where $\mathbf{Z}$ represents the network's n_Z EMs and $\mathbf{r_M}$ represents the regulated uptake rate of each EM. Given the use of the pseudo-steady state hypothesis, the changes in extracellular concentrations are tracked in the following way

$$\frac{1}{c}\frac{d\mathbf{x}}{dt} = \mathbf{S_x Z^T r_M} \tag{3.11}$$

As stated previously, in kinetic models, the reaction rates of each chemical transformation are tracked. In HCM, the regulated rate expressions for $\mathbf{r_M}$ take similar Michaelis-Menten forms, however, they include regulation, v_i and enzyme e_i terms

$$r_{M,i} = e_i v_i \frac{r_{M,i}^{\text{kin,max}} s_i}{K_{M,i} + s_i} \tag{3.12}$$

Above, parameters $r_{M,i}^{\text{kin,max}}$ and $K_{M,i}$ are similar to the parameters from the kinetic model with one main distinction. They describe the rate of uptake into an EM or set of reactions instead of the rate of a single reaction. In HCM, the change of enzymes which regulates $\mathbf{r_M}$ is

$$\frac{de_i}{dt} = \alpha + u_i \frac{k_{E,i} s_i}{K_i' + s_i} - (\mu + \beta) e_i \tag{3.13}$$

In HCM, enzymes are generated by a constitutive formation and induced formation which make up the first two terms above. The last is an expression for the depletion of enzymes due to growth dilution and degradation. Cybernetic control variables u_i and v_i guide the induced synthesis of enzymes and the allosteric regulation of enzyme activity respectively. Induced enzyme formation is expressed as some function of each pathway's return on investment (ROI), p_i. ROIs typically are defined as each pathway's rate of substrate uptake or growth rate. To calculate the control of enzyme formation, ROIs are compared for each pathway in

$$u_i = \frac{p_i}{\sum_j p_j}, \tag{3.14}$$

where the denominator represents the sum of ROIs for all pathways. This means that the fraction of a finite resource pool devoted to the production of enzymes for one

pathway is proportional to the ROI for that pathway. Similarly, the activity of the different metabolic pathways is controlled by the v_i variable which takes the form

$$v_i = \frac{p_i}{\max_j p_j}. \tag{3.15}$$

The pathway with the highest ROI will be fully expressed. All other pathways with lower ROIs will be down-regulated proportionally.

For a given metabolic network, the number of elementary modes can be quite high. Therefore, yield analysis is used to reduce the elementary modes in HCM down to a minimal set that spans a given yield space [22]. This makes the generation of an HCM model for the subsequent results facile and reduces the number of parameters for the model's specification.

Another version of cybernetic models that will be analyzed is the lumped hybrid cybernetic model or L-HCM. The formulation of this model is quite similar to HCM with one main distinction. Instead of enumerating uptake rate constants for all pathways, the EMs are lumped together into families based on their structural returns on investment. These family modes are then expressed as a function of some dynamic metabolic objective function. The procedure for lumping EMs is somewhat complex and is best explained in [24].

Both HCM and L-HCM employ objective functions to dynamically maximize the rate of carbon uptake in the models used in the subsequent scenarios.

3.2.3 Comparison Method

To compare the ability of each model with one another, both AIC and BIC are computed for the set of models to gauge which one minimizes their values. Consistency among the modeling frameworks for minimizing both criteria are considered. The metabolic model that best minimizes these information criteria for different scenarios is identified as the best model for the purposes of optimization and control applications. In other modeling scenarios, such as those in which models are developed for the purposes of biological discovery, AIC and BIC should not be used for

model selection as quantifying the tradeoff between accuracy and simplicity is not necessarily relevant.

3.3 Results

3.3.1 Modeling Dynamic Fluxes of the Aerobic Growth of *E. coli*

To compare how well various biological models compress metabolic fluxes, four different models describing the aerobic growth of *E. coli* were compared. In this growth scenario, *E. coli* first consumes extracellular glucose while generating acetate as a byproduct. Once the glucose has been exhausted, the culture then shifts its metabolic state to consume the acetate product. The models chosen to describe various aspects of this system are (1) a detailed kinetic model, (2) dynamic Flux Balance Analysis, (3) HCM, and (4) L-HCM.

The reaction network used in this model selection exercise was taken from Kotte et al. [29] whose metabolic network is summarized in Figure 2 which was also used as the kinetic model for this comparison. The network includes a variety of metabolic pathways including glycolysis, gluconeogenesis, and the TCA cycle. Note that some metabolic reactions are truncated with others in order to simplify the network. For example the network's reaction for G6P's conversion to FBP lumps together two reactions and ignores the intermediate product of F6P.

The kinetic model [29] is taken as the basis for this exercise as this model type is generally the most labor intensive to develop in terms of parameters and structure. Then, dFBA, HCM and L-HCM formulations were developed for the kinetic model's growth scenarios. The structured biomass equation used to develop these models was extrapolated from a prior *E. coli* model [57] and takes the form of

$$\begin{aligned} 3.7478\ \mathrm{AcCoA} + 5.1971\ \mathrm{alKG} + 1.32915\ \mathrm{G6P} \\ +18.271\ \mathrm{NADPH} + 1.787\ \mathrm{OxA} + 0.5191\ \mathrm{PEP} \\ +1.496\ \mathrm{PG3} + 2.8328\ \mathrm{Pyr} + 59.89551887\ \mathrm{ATP} \\ \rightarrow 1\ \mathrm{Biomass} + 3.547\ \mathrm{NADH} \end{aligned} \tag{3.16}$$

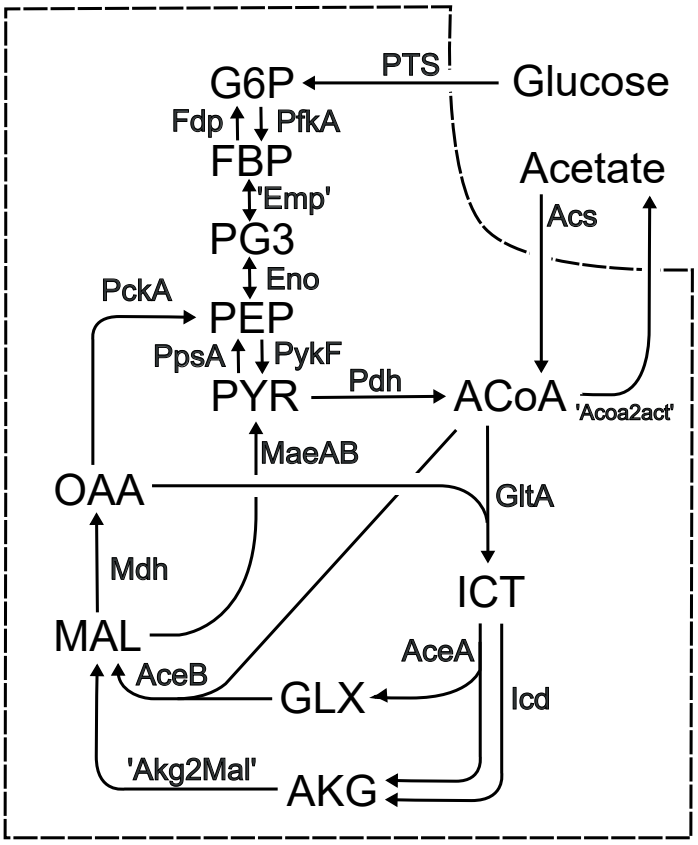

Figure 3.2.: Schematic of the simplified network used to construct the various models. Reaction names used in subsequent plots are listed next to their respective arrows.

3.3.1.1 Steady State Flux Predictions

Each of the models tested generates a description of this network's metabolic fluxes for both dynamic scenarios and steady state scenarios. Each one of these model descriptions of fluxes can then be verified using steady state flux data for growth on glucose and acetate from [58]. Figure 3 shows the correlation plots describing the accuracy of each model's steady state flux description with the appropriate Pearson's product-moment correlation coefficient. Note that FBA is used to calculate the steady state fluxes instead of dFBA and that the kinetic model was parameterized, in part upon the flux data.

Figure 3 clearly shows the kinetic model's superior ability to describe the rates of metabolic reactions. Nonetheless, it is evident that all models provide reasonable descriptions of metabolic fluxes. The Pearson metric is listed in table I for each condition. The model with the most parameters, being the kinetic model with 193 experimental constants, produced the most accurate approximation of the experimental flux values for both growth on glucose and acetate. The next most accurate experimental flux description is produced by L-HCM. Despite ranking third and fourth, HCM and FBA also provide good approximations of the experimental fluxes of growth on glucose. Both HCM and dFBA, however show low accuracy in generating steady state approximations of the acetate fluxes.

Model comparison metrics were calculated for both the glucose and acetate conditions to gauge how well each model performed. Parameters for the models were tabulated. L-HCM was the second most parameterized model with 14 parameters. These parameters include the kinetic parameters as well as those used to lump the elementary modes together. HCM had 12 parameters. FBA had only 3 parameters which were the uptake and excretion rates of substrates and products in this system. Note that the objective functions in these different models were not parameters, but structures for the various models which did not imply additional penalty for model complexity.

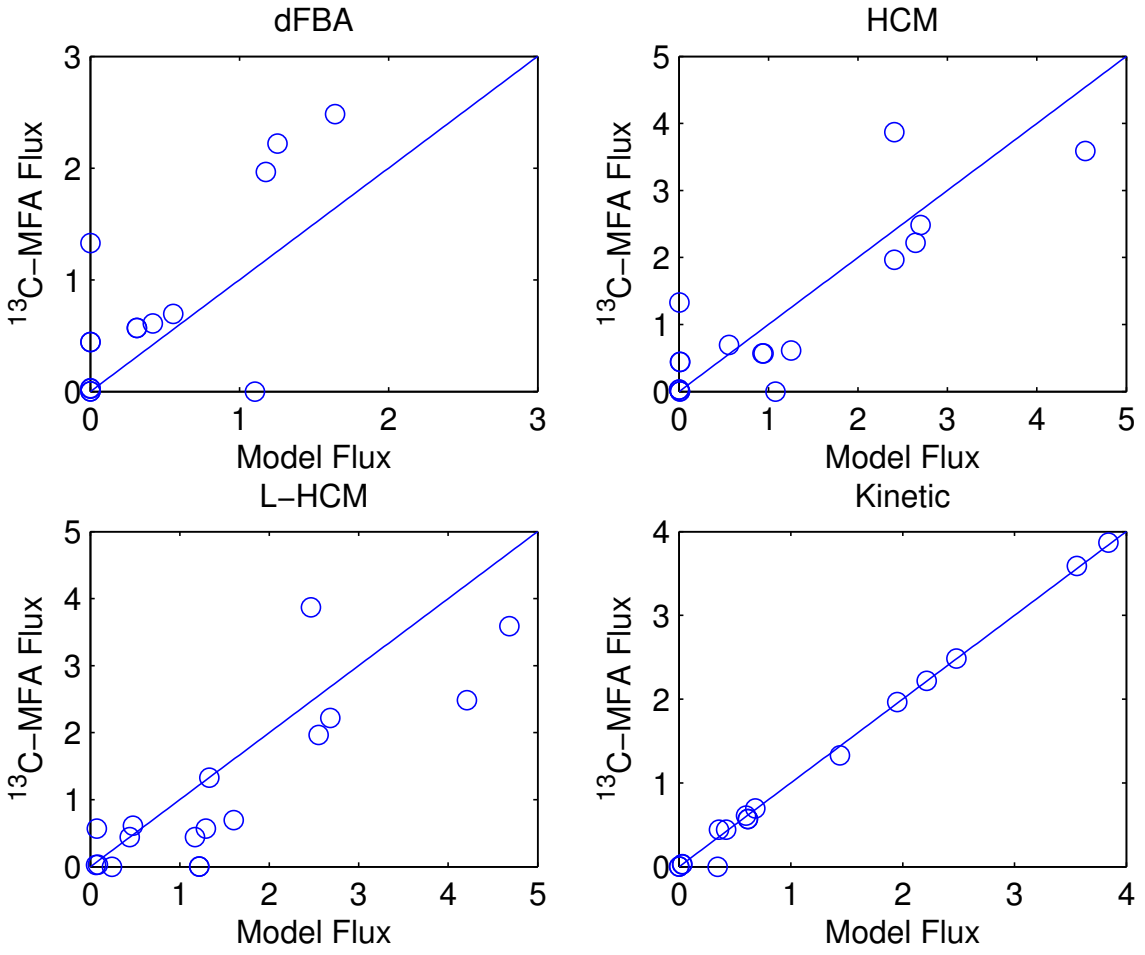

Figure 3.3.: Correlation plots for each model's description of steady state flux data for *E. coli* growing on glucose. Horizontal axis shows the flux values predicted by each model and the vertical axis corresponds to the experimental values for steady state fluxes taken using carbon-13 labeling experiments.

Table 3.1.: Correlations, parameterization and information criterion for the model set for the steady state fluxes.

	Kinetic	FBA	HCM	L-HCM
Parameter No.	193	3	12	11
$\rho_{Glucose}$	0.9972	0.8172	0.8269	0.8548
$AIC_{Glucose}$	113.38	**1.1445**	4.0746	3.6997
$BIC_{Glucose}$	193.78	**2.3943**	9.0739	8.2823
$\rho_{Acetate}$	0.9980	0.3290	0.3307	0.7858
$AIC_{Acetate}$	186.63	14.411	18.635	**0.3933**
$BIC_{Acetate}$	267.03	15.661	23.634	**4.9760**

The FBA model for the glucose steady state fluxes showed the best minimization of both information criterion. FBA, which demonstrated poor performance in predicting the fluxes for the steady state growth on acetate and was outperformed by L-HCM for both AIC and BIC. Despite the kinetic model's very high correlation with experimental flux data for both conditions, it was severely penalized by its large number of parameters.

3.3.1.2 Dynamic Flux Predictions

Metabolic fluxes are difficult to measure experimentally and require carefully controlled experiments. Because of this, time-series experimental fluxes are unavailable for this system. To compare the dFBA, HCM and L-HCM models on their ability to model dynamic fluxes, the kinetic model's prediction of dynamic fluxes will be treated as an experimental approximation of the true dynamic metabolic fluxes for this system. The use of this artificial data is justified in part by the fact that the model was parameterized upon a wide variety of data for multiple levels of cellular processes. It incorporates a great span of regulatory phenomena including transcription factors,

transcription-factor metabolite interactions, gene expression, enzyme production, kinase reactions, phosphatase reactions and protein degradation. Given the high degree of complexity of the kinetic model, it will be reasonable to assume that it provides a close to true approximation of the real dynamic fluxes for the *E. coli* system that is being modeled.

Figure 4 shows the comparison of the information theoretic metrics with the artificial data. Generally speaking, HCM and L-HCM overpredict the flux rates through the glyoxylate pathway. They also overpredict the flux of malate to pyruvate during growth on both glucose and acetate. On the other hand, dFBA underpredicts the fluxes through glycolysis and the TCA cycle while L-HCM and HCM provide good qualitative descriptions of the simulated data. L-HCM overpredicts the futile cycling from G6P to FBP during the consumption of glucose while HCM and dFBA underpredict this. HCM best predicts the flux through pyruvate dehydrogenase while dFBA is lower and L-HCM is higher.

As in the steady-state flux model comparison, model selection metrics were calculated for the set of models. The complexities of the HCM and L-HCM models did not change. However, to incorporate the dynamics of the system, 5 additional parameters were added to generate the dFBA description of the data. The artifical data was generated at 15 min intervals for a 10 hour period of growth. The model that best minimizes the information criterion for the dynamic flux data is HCM which is shown in Table II. This is followed by L-HCM. dFBA places last for both information criterion.

3.3.2 Another Case Study: FBA vs. L-HCM in *S. oneidensis* Flux Prediction

Case Study: A Comparison of L-HCM and FBA For another illustration of IC and their ability to distinguish models, these concepts can be used to compare FBA and L-HCM for flux predictions of Shewanella oneidensis [26]. Starting from the flux data [59] as a means for testing the accuracy of model predictions, AIC and BIC values were calculated for each respective model. Note that the range of predictions

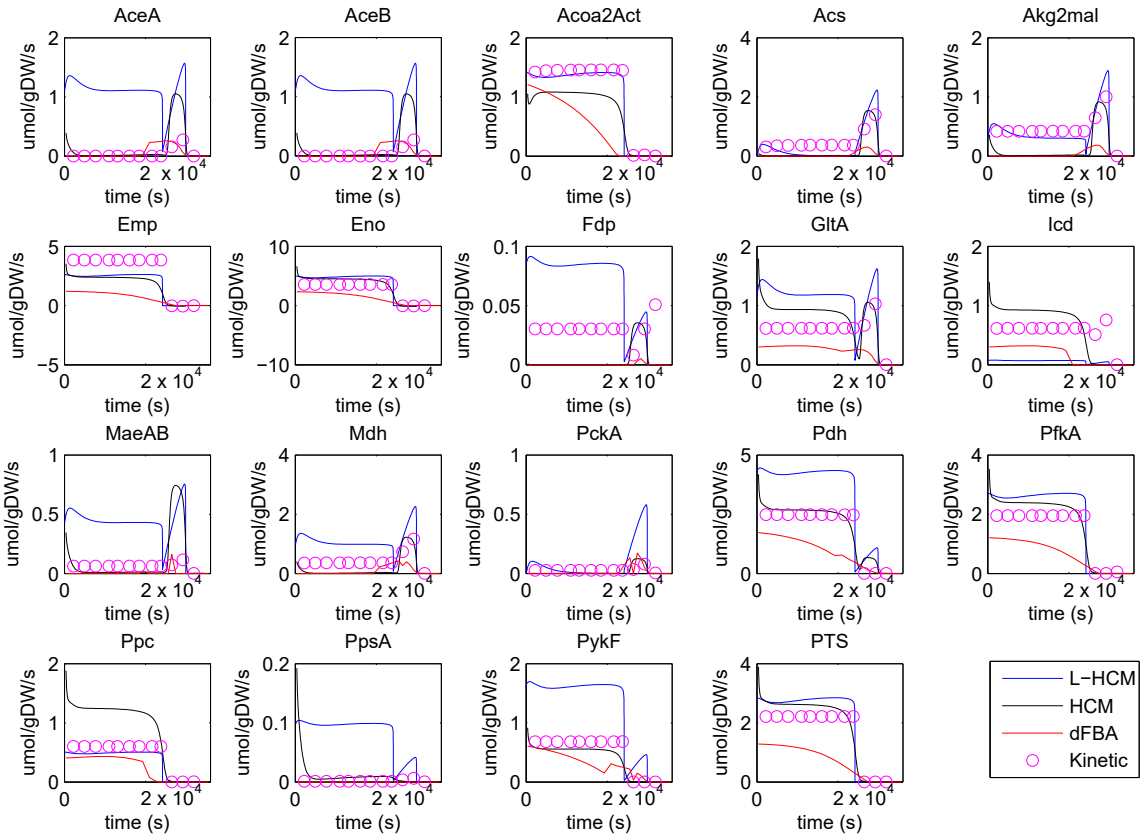

Figure 3.4.: Dynamic flux profiles for all three models. The flux titles correspond to the reaction names given in Figure 2. The purple circles represent the values for the artificial flux data. The red line represents flux predictions made by dFBA. the black and blue lines represent the flux predictions made by HCM and L-HCM respectively.

Table 3.2.: Information criterion for dynamic flux descriptions for the artificial data.

	FBA	HCM	L-HCM
Parameter No.	7	12	11
AIC	3580.8	**-1607.4**	2926.5
BIC	3612.7	**-1552.7**	2976.6

is due to variability in interpretation as to how a parameter is defined. That is to say, for instance, in the case of L-HCM, only (3-6) parameters are trained upon data. The remaining (10 or so) parameters are generally insensitive ones. Refer to figure 3.5 in order to see how AIC and BIC values compare for the various model classes. From this table, the conclusion that L-HCM delivers, in relative terms to FBA, accuracy at a low cost of model complexity is drawn.

3.4 Discussion

The results provided by the model selection framework presented in this work varied by application. In the comparison of the different models' abilities to predict steady state flux data, there was a mixed outcome. While FBA was highly capable of describing the steady state fluxes for the growth on glucose, it was less able to provide an accurate description of the fluxes for the growth on acetate. L-HCM, conversely, placed third in its minimization of the information criterion for glucose. However, it best minimized them for the acetate fluxes. L-HCM also provided the best steady state flux description for the glucose fluxes, after the kinetic model, but its complexity penalized it into the third place. This additional complexity, allowed it to capture the acetate steady state fluxes best. Because of the information criterion's ability to penalize models for their level of complexity, FBA has a clear advantage when it is able to provide an accurate estimate.

The ability of L-HCM to capture steady state fluxes has also been demonstrated in prior work [26]. The strength of L-HCM is that it combines multiple elementary modes into a smaller subset of lumped elemantary modes. The lumping procedure takes into account experimental data related to the yield of products and biomass for various substrates. The strength of this lumping procedure is clearly seen in the steady state flux results.

When treating these models as data compressing entities, it is noteworthy how much FBA, HCM and L-HCM reduce the values of the information criteria relative to the complex, kinetic description of the steady state fluxes. All three models reduce

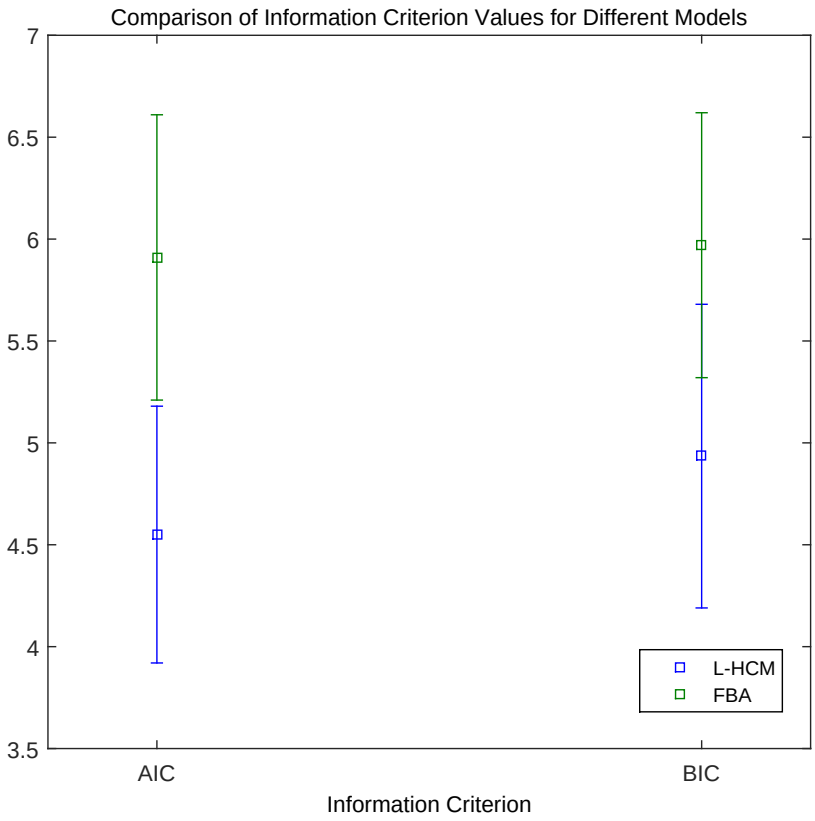

Figure 3.5.: Information criterion for comparison of FBA (green) and L-HCM (blue) flux predictions. Data point normalized AIC and BIC values are on vertical axis and grouped on horizontal axis. Lower IC scores imply more compact description of data.

the information criterion values by 2 orders of magnitude representing a significantly reduced description length for the data. It is intuitive that the complex kinetic description of the system would provide the most excessively complex description of the equilibriated flux state for this system. The information criterion are able to take this intuitive statement and establish it in quantitative sense. Moreover, they show the degree to which the kinetic model over-explains steady state data.

The scenarios tested above showed that HCM provides the best description of the artificial dynamic flux data despite the fact that it did not perform well in estimating the steady state fluxes. This is most likely due to the fact that HCM employs a combination of elementary modes over time to describe the changes in fluxes. In steady state scenarios, HCM will only express one of these numerous elementary modes that embodies some extreme edge of the yield space as determined using yield analysis. This comes from the cybernetic policy which will selectively produce only one pathway's bulk enzyme pathway once a steady state is reached because there is only one pathway that will have a maximum unregulated rate for the substrate concentration present at a steady state. Notwithstanding HCM's difficulty in the capture of steady states for both of the substrate conditions modeled here, the inclusion of multiple elementary modes for a single substrate becomes an advantage for HCM as it can describe the span of an organism's metabolic yield space.

In contrast to HCM, the L-HCM model shows a markedly less accurate prediction of the dynamic fluxes for this system. Contrasting accurate steady state predictions, it is less capable of reproducing the artificial data for dynamic fluxes in this scenario. This might in part be due to L-HCM overpredicting the overall flux through different pathways which stems from how it lumps together the different modes. Also, it reduces the complexity of the model into merely two elementary modes which may not be sufficient in spanning the total space of allowable fluxes. This could mean that despite its dynamic objective function, it is limited to a more limited space in its description of dynamically altering flux profiles. Regardless, both HCM and L-HCM outperform dFBA in the compression of dynamic flux data. This makes

the argument that rate-based objective functions are more descriptive of the data compared to yield-based ones.

On the whole, dFBA provides the least accurate prediction of dynamic metabolic fluxes. This could stem from the inadequacy of a yield-based objective function in describing the dynamic shifts in metabolic states. dFBA consistently underpredicted the values of fluxes which may be due to the biomass maximization objective function. This is due to the formulation of the LP problem where maintenance and other functions related to metabolism are ignored and therefore the total sum of fluxes may be contracted from a realistic state. In other words, the goal of the FBA model is to convert the most substrate into biomass with a given set of constraints and will ignore phenomena such as futile cycling. Other objective functions such as total flux minimization could provide better results but were not tested in this work. It has been shown that the accuracy of objective functions vary by scenario [48]. This would also help to explain the biomass maximization objective functions differences in accuracy when describing the steady state scenarios. Also not tested were additional constraints on flux values for the metabolic system.

The artificial data for the fluxes was a reasonable substitution given the lack of actual dynamic data for this model system. It is possible that dynamic flux data may be different from what is indicated by the kinetic model. However, given the fact that the kinetic model incorporates a great number of regulatory phenomena and captures steady state fluxes with near-perfect accuracy, it is likely the best possible approximation of the dynamic fluxes.

The fact that the model selection criterion were minimized for each model for a different application makes clear the fact that certain models are more relevant to compress specific types of data. It is natural that FBA would be able to minimize these information criterion for the steady state scenarios given its low complexity. In the acetate steady state case, however, the fact that L-HCM outperforms FBA brings to light the fact that FBA's objective functions are not universally descriptive nor applicable in all substrate consumption scenarios. HCM, with its incorporation

of multiple elementary modes finds a balance between these two which captures the dynamic states predicted by the kinetic model best.

Also, the varied results by scenario point out the fact that these information criteria are not biased towards one approach. Their ability to compare vastly different formulations for the same metabolic system highlights their utility in systems biological applications. Information criterion treat each model as an alternate description length for the data. A key assumption of these metrics is that the model structure itself is not communicated, only the parameters. This assumption allows these metrics to compare non-nested models.

Finally, both AIC and BIC were minimized for the same model for all three model selection scenarios. This consistency is a good indication of their utility in reaching objective conclusions for model selection.

3.5 Conclusion

This work has shown, for the first time, how information criterion can be used for the comparison of non-nested systems biological models. While these metrics have been well-established for many applications, their use has not been brought to the attention of metabolic modelers who could benefit from a deeper understanding of how different models balance between accuracy and complexity. This work has shown that L-HCM provides the most succinct description of steady state fluxes for *E. coli* growing on acetate. It has also demonstrated that FBA optimizes the information criterion for *E. coli* growing on glucose at steaty state. Finally, it has shown that HCM minimizes these metrics for the description of artificial dynamic flux data.

These conclusions, however, are contingent upon the objectives of the modeling effort. Application of models for the purposes of optimization and/or control requires compromising model complexity in favor of rapid assessment of predictions. On the other hand, a thorough understanding of the changes in metabolic performance due to engineered perturbations can only come about by using detailed models without serious compromise of complexity. Our analysis demonstrates quantitatively how dif-

ferent modeling frameworks perform when the model objectives are defined. Thus conclusions made for one objective will not necessarily carry over to other circumstances.

4. INVESTIGATING THE RELATIONSHIP BETWEEN CYBERNETIC VARIABLES AND TRANSCRIPTOMIC DATA: DIAUXIC GROWTH

Summary

Diauxic growth of *E. coli* is driven by a host of internal, complex regulatory actions. In this classic scenario of cellular control, *E. coli* prioritizes the utilization of the substitutable substrates present in its external environment. Cybernetic models of metabolism, whose development now spans three decades, were first formulated to describe the growth behavior of cells in such multi-substrate environments. These models utilize the hypothesis that the synthesis and activity of the enzymatic machinery is regulated to maximize a return on investment. While this assumption is made on the basis of logical arguments rooted in evolutionary principles, little effort has been spent to test the extent to which this reasoning mimics the regulation of enzyme synthesis and activity in microorganisms. This work revisits the cybernetic model describing diauxic growth and compares its predictions of enzyme synthesis control with time series gene expression data in microarray and RT-PCR formats. Three separate studies are made for two different strains of *E. coli*. The first study is for the growth of *E. coli* BW25113 on a mixture of glucose and acetate whose gene expression changes are monitored using microarray. The second study is for the growth of *E. coli* MG1655 on glucose and acetate whose expression changes are monitored using RT-PCR. The final study is for the growth of *E. coli* MG1655 on glucose and lactose where gene expression changes are monitored using microarrays. By demonstrating that the cybernetic variables for induced enzyme synthesis mimic changes in transcriptional data, we show that the cybernetic models capture not only the dynamic trends in the concentrations of extracellular substrates and products but also accurately approximate the regulatory phenomena in microorganisms.

4.1 Introduction

Cybernetic models, which describe the dynamic regulation of metabolism as a goal-directed process, have been developed over the past three decades to describe a host of complex microbial growth scenarios. The first scenario cybernetic models sought to describe was that of diauxic growth of microorganisms [3]. These models assume that microorganisms are optimal control systems that actively modulate their metabolism through the cybernetic variables for enzyme synthesis, u_i, and activity, v_i. These models further assume that microorganisms possess limited internal resources that they use in an optimal manner to achieve their goals. Using matching and proportional laws to describe enzyme synthesis and activity respectively, these first cybernetic models were able to predict the distinct growth and lag phases of the diauxie using parameters taken from growth data on individual substrates.

From this initial formulation, cybernetic models have been extended to describe much more complex aspects of metabolism via incremental improvements over the years. Cybernetic models have been useful in not only describing complex substrate uptake patterns [9], but have also yielded successful predictions of intracellular fluxes [26], gene-knockout behaviors [25], and multiplicity of steady states in chemostats [21]. While cybernetic modeling's ability to predict complex cellular phenomena helps to realize the utility of this approach, no effort has taken place to verify the extent to which the cybernetic control mechanisms mimic cellular regulation. More specifically, the cybernetic variables for enzyme synthesis and activity, u_i and v_i, have not been directly compared with cellular data that is representative of the regulatory mechanisms in cells.

Experimental advances in the past decade have resulted in the collection of large amounts of cellular data which provide insights into regulation in cells [60,61]. Various techniques such as microarrays and RT-PCR are used to monitor the expression of mRNA from a large number of genes in cells under a variety of experimental conditions. These techniques have been used to characterize the gene expression

changes that occur when *E. coli* switches from growing on one substrate to growing on another.

To briefly explain the changes that occur, when *E. coli* switches from growing on glucose to growing on acetate, a number of regulatory mechanisms are activated that monitor the cell's nutritional status and the propensity for internal acidification [62]. These mechanisms then coordinate a host of actions on the gene expression level to induce the uptake of acetate predominantly through the ACS pathway where it is converted to other growth precursors via increased flux through the TCA cycle [63]. Glycolysis ceases and carbon is channeled in a completely reversed direction via gluconeogenesis. Similarly, when *E. coli* switches from consuming glucose to consuming lactose, a more limited number of events occur. As glucose is depleted, cAMP levels are elevated and crp dependent factors promote the expression of the lac operon which has already been derepressed by the presence of external lactose [64]. The *lac* operon is then transcribed and translated into enzymes relevant to the metabolism of lactose.

The present chapter focuses on comparing the predictions for cybernetic control variable for enzyme synthesis ui with data that is representative enzyme synthesis available in the form of mRNA measurements and measured using microarrays and RT-PCR. It should be noted that these control variables are not a direct substitute for gene expression and that the comparison made focuses on the relative changes and not absolute level of mRNA. However, given the fact that these cybernetic control variables describe the induction of enzyme synthesis, they should change in a temporal fashion that is consistent with changes in mRNA levels.

Three separate studies wherein microorganisms grow on two substitutable substrate pairs are made. The first study is for the growth of *E. coli* BW25113 on a mixture of glucose and acetate whose gene expression changes are monitored using microarray. The second study is for the growth of *E. coli* MG1655 on glucose and acetate whose expression changes are monitored using RT-PCR. The final study is for the growth of *E. coli* MG1655 on glucose and lactose where gene expression changes

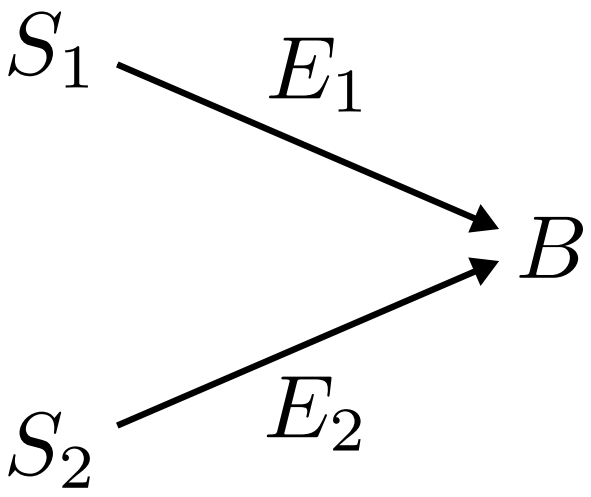

Figure 4.1.: A simple schematic of the two competing metabolic pathways during diauxic growth.

are monitored using microarrays. By demonstrating that the cybernetic variables for induced enzyme synthesis mimic transcriptional data, we show that the cybernetic models capture not only the dynamic trends in the concentrations of extracellular substrates and products but also produce a control strategy that is consistent with cellular data.

4.2 Materials and Methods

To describe the diauxic growth phenomena that occur in the two scenarios introduced, cybernetic models were developed. These models determine the cybernetic control of each lumped pathways enzyme synthesis and are formulated using the structure of prior cybernetic models [3, 4]. A schematic of this lumped type of model is in figure 4.1.

Most simply, a cybernetic model of metabolism is a set of coupled, first-order ODEs that describes the time-dependent rates of change in the concentrations of cells, metabolites and enzymes. For each substrate, S_i, there is a lumped enzyme set

E_i that is employed to digest each substrate and a yield describing the total amount of biomass, B, produced from each substrate according to the reaction:

$$B + S_i \xrightarrow{E_i} (1 + Y_i)B + ... \tag{4.1}$$

In diauxic growth, once the cell consumes the entirety of the preferred substrate, the production of lumped enzyme for the secondary substrate commences. In the cybernetic formulation, the rate at which each substrate is converted into biomass is governed by the concentration of the amount of pathway specific enzyme, e_i, present in the system. Also influencing this rate of reaction are the enzyme specific growth rate μ_i, substrate concentration, s_i, and the total concentration of cells c. This rate is formulated as

$$r_i = \frac{\mu_i e_i s_i c}{K_i + s_i}, \tag{4.2}$$

where K_i is the substrate specific Michaelis-Menten constant. The formation of biomass is determined by the summation of the activity controlled biomass formation reactions and is written:

$$\frac{dc}{dt} = \sum_i r_i v_i. \tag{4.3}$$

Each substrate's depletion is given by the negative substrate specific biomass formation rate normalized by the substrate's yield.

$$\frac{ds_i}{dt} = \frac{-1}{Y_i} r_i v_i \tag{4.4}$$

Finally, the production of the lumped enzymes is as follows

$$\frac{de_i}{dt} = \alpha + u_i \frac{k_{E,i} s_i}{K_i' + s_i} - (\frac{d \ln c}{dt} + \beta) e_i. \tag{4.5}$$

The above equation includes a number of terms for the phenomena related to changes in enzyme concentration. The first is for the constitutive enzyme synthesis, α [4]. The second is for induced enzyme synthesis which is controlled by the cybernetic variable, u_i. The rate of enzyme formation is established by a rate constant for induced enzyme

synthesis $k_{E,i}$, and is a function of the substrate level and a Michaelis-Menten constant K_i' . Lastly, the enzyme quantity is reduced both by dilution due to growth and from enzymatic degradation β. Note that this enzyme control variable u_i appearing above will be the model variable that is compared to gene expression data.

4.2.1 Cybernetic Control Formulation

In the above model formulation, two control variables are presented. Control variable u_i specifies the regulation of enzyme synthesis for different pathways and v_i specifies activity adjustments of the lumped pathways. To efficiently allocate cellular resources, single cells must make decisions regarding the regulation of competing metabolic pathways. From this model's perspective on diauxic growth, there are two competing metabolic options for which the cell can invest resources in to grow. At any given time, there is an unregulated growth rate for each substrate given by r_i which is a function of the amount of lumped enzyme present. This enzyme quantity, in turn, changes as a function of the instantaneous rate through the u_i control variable's effects. This control variable compares the different metabolic options and invests resources proportionally to a metabolic pathway's return on investment growth and is written as the following matching law:

$$u_i = \frac{r_i}{\sum_j r_j} \tag{4.6}$$

In this simple cybernetic model, return on investment for a lumped metabolic pathway is equal to its instantaneous rate of growth. This implies that cells prioritize their limited capacity for enzyme synthesis for pathways that have a higher payoff. This variable is normalized by the sum of returns on investment which represents the total pool of resources available to convert sugars into biomass. The set of u_i control variables sum to 1. In this control variable, it is evident that there will be a higher investment of resources into the lumped enzyme for a pathway if there is a higher the return on investment or rate of growth.

In addition to the efficient allocation of resources, pathways are also regulated by mechanisms that adjust their activity. To model this, pathways that have lower rates of growth than the maximal pathway are turned down via the proportional law which is

$$v_i = \frac{r_i}{\max_j r_j}. \tag{4.7}$$

This is justified in part by the fact that some metabolic pathways may require fluxes to route through different parts of the metabolic network in opposite directions. For example, in the case of glycolytic and gluconeogenic substrates, glucose and acetate, pathways for one or the other should be turned down in order to reduce futile cycling. Therefore, the proportional cybernetic variables turn down pathways that do not represent the highest return on investment and fully express the pathway that does.

4.2.2 Estimation of Parameters

To develop two of the models in this work, sets of data including both batch growth data and time series microarray data taken during the switching of substrates were used. One data set is from [65] describing the growth of *E. coli* BW25113 on a mixture of glucose and acetate. The other set from [66] elucidates the changes in gene expression for *E. coli* MG1655 growing on a mixture of glucose and lactose. For both of these data sets, a growth curve and initial substrate data are available. Given that microarray data is inherently noisy, the findings for the glucose-acetate growth scenario were verified with RT-PCR data. An additional model was developed to describe the *E. coli* MG1655 strain using data from [67]. RT-PCR data was not available in the literature to further validate the result from the glucose-lactose diauxic growth scenario.

Models were first parameterized for the microarray data. From the growth curves in the glucose-lactose condition, the parameters μ_i^{max} and Y_i were estimated from the knowledge of the initial substrate levels, biomass concentrations and slopes of the growth curves. Michaelis-Menten parameters were at first taken from [3]. Michaelis-

Table 4.1.: Parameter values for diauxic models of *E. coli*. Strains MG1655 and BW25113 growing on multiple substrate sources.

Strain	MG1655		BW25113	
Parameter (units)	Glucose	Lactose	Glucose	Acetate
$S_{i,0}$ (g/L)	0.5	1.5	1.5	0.9
μ_i^{max} (h^{-1})	0.642	0.618	0.442	0.426
Y_i (gDW/g)	0.400	0.440	0.339	0.129
K_i/K_i' (g/L)	0.005	0.50	0.0088	0.3317
α_i	1.67e-3	1.67e-3	1.67e-3	1.67e-3
$k_{E,i}$	1e-3	1e-3	1e-3	1e-3
β_i	0.05	0.05	0.05	0.05

Menten constants, yields and growth rates for the glucose-acetate data were determined using available single substrate data. Once these initial values were collected for both models, a genetic algorithm was used to refine the initial parameter values using a normalized least squares error function for the model's fit of the biomass data. The parameter values are listed in table I. Note that the Michaelis-Menten parameters K_i and K_i' take on the same value for both r_i and the induced enzyme formation terms for the same lumped metabolic pathway. For example, K_i and K_i' have the same value of 0.05 for the glucose growth pathway describing the glucose-lactose diauxie of *E. coli* MG1655.

The RT-PCR model was built in a similar fashion. The main departure in its formulation from the other models is the inclusion of an acetate production term. In this scenario, glucose is the only substrate initially. While consuming glucose, acetate is produced via overflow metabolism to adjust the intracellular redox ratio [68] which is then consumed after glucose is exhausted. Note that acetate production is absent from the microarray models as the dynamic changes in substrates and products

are not given in the datasets and modeling this phenomena would require extraneous assumptions. For more details on the parameterization and development of this particular model, refer to the appendices.

4.2.3 Comparison of Dynamic Gene Expression Data and Model Variables

The general goal in the comparison between the u_i control variables and dynamic gene expression is the demonstration of qualitative similarity. These two entities cannot be compared in an absolute sense because their magnitudes do not align. The matching law variables are constrained between values of 0 and 1 while gene expression levels vary over a greater range depending on how they are normalized. Moreover, given the nature of microarray, it would not be useful to make absolute comparisons as it is often referred to as a "semi-quantitative" method. Therefore, it is pertinent to normalize both the control variables and gene expression in a fashion that complements qualitative comparison. The method elected to do this is a standard normalization procedure where the data is scaled by the standard deviation and centered around zero by subtracting a series mean [69]. For some series $C = \{c_1, c_2, ...c_n\}$ of either gene data or model variables, the normalization is

$$c_i' = \frac{c_i - \mu(C)}{\sigma(C)}, \tag{4.8}$$

of which $\mu(C)$ is the mean of the data series and $\sigma(C)$ is the standard deviation of the series. The mean of a data series $\mu(C)$ is not to be confused with the enzyme specific growth rate for a given pathway, μ_i. This normalization method aligns the data in such a way to make the comparison of the dynamic trends in the data more clear. To quantify similarity between the model variables and experimental data, the use of correlational statistics such as Pearson's correlation coefficient is still appropriate as this statistic is scale invariant. The microarray and RT-PCR data were already provided as gene expression normalized sets in their original sources [65–67]. These data sets were not processed in any additional way.

4.3 Results

4.3.1 Model Capture of Biomass Formation

The models developed were able to accurately describe the growth of biomass. All systems modeled capture the initial growth phase on glucose and the second growth phase on either lactose or acetate. They also correctly represent the timing of the lag phase as the batch culture transitions from the preferred substrate to the secondary one. The predictions of these models is visualized in figure 4.2.

From these model descriptions of the data, it can be gauged how the values of the control variables for each pathway change dynamically. As the glucose level depletes, the u_i variable for the glucose pathway decreases from one towards zero. This transition is made during the lag phase of the diauxic growth. At the same time, the u_i variable for the lactose or acetate pathways increase from a value close to zero towards one. It is at this critical point that the comparisons between the control variable and gene expression will be made.

4.3.2 Study of Glucose-Acetate Diauxie

E. coli growing in a mixture of glucose and acetate can employ either glycolytic and gluconeogenic pathways to synthesize biomass and grow. In this mixture, the culture first consumes glucose and switches to acetate, a carbon source upon which it has a lower growth rate. When the culture shifts from the digestion of glucose to acetate, glycolysis is unneeded and gluconeogenic reactions must channel carbon to synthesize essential biomass precursors. A host of other changes happen in the central carbon metabolic pathways. Intuitively, acetate secretion goes down and uptake increases. Fluxes through the glyoxylate pathway and TCA cycle increase [70]. Also, the fluxes through the pentose phosphate pathway are significantly higher during growth on glucose compared to acetate.

These changes in the various metabolic pathways are observable on the gene expression level. The cybernetic variable profiles have been compared accordingly with

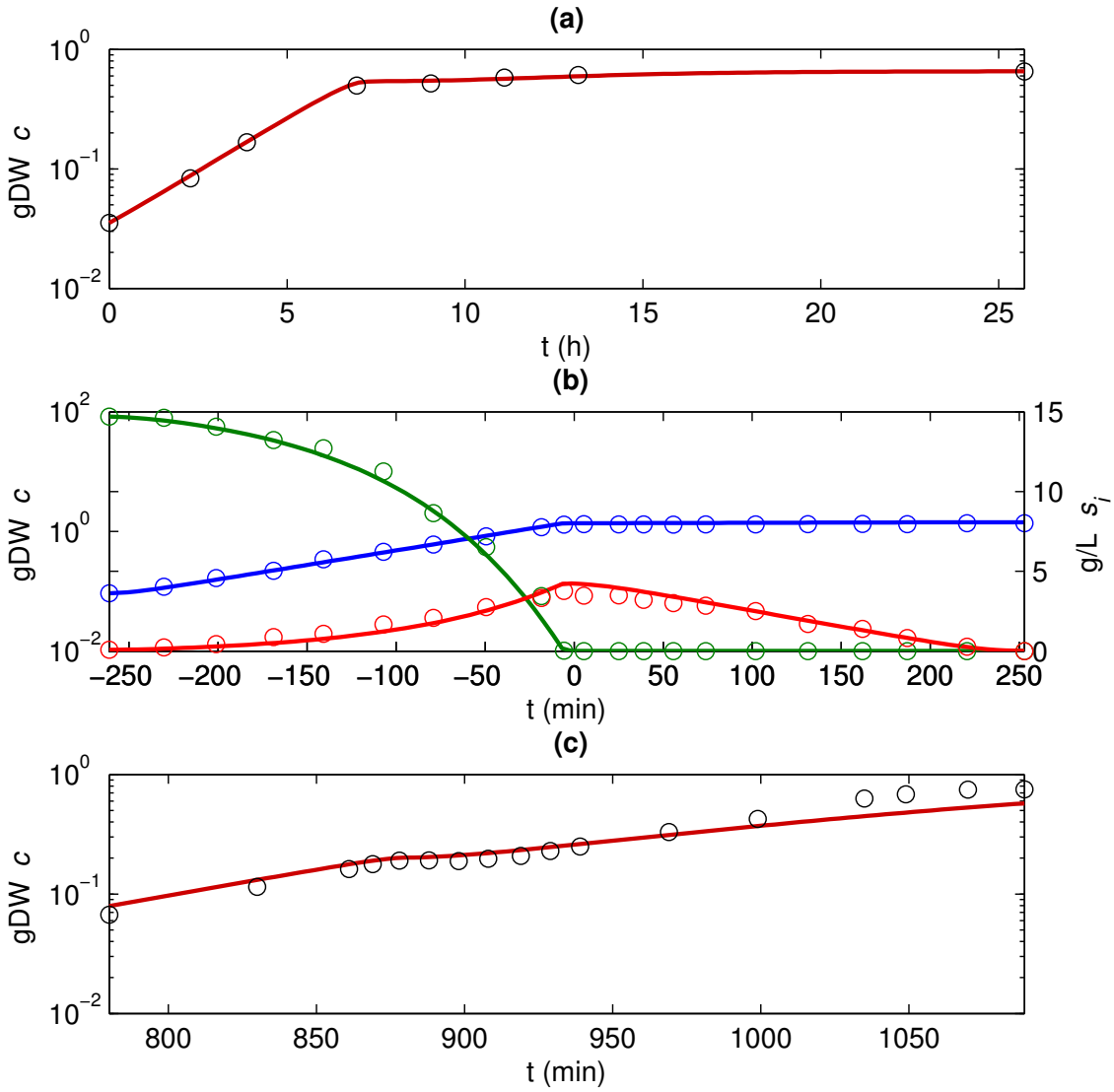

Figure 4.2.: Predictions of external metabolite concentration data for the three diauxic systems. (a) shows the biomass level for *E. coli* BW25113's diauxic growth on glucose and acetate. (b) exhibits *E. coli* MG1655's growth on glucose and acetate where RT-PCR data are available. The left axis shows biomass, c, in gDW. Substrate concentrations, s_i, are in g/L and are tracked on the right axis with the time profile of glucose in green and acetate in red. (c) shows diauxic growth of *E. coli* MG1655 on glucose and acetate.

the dynamics of these various central carbon metabolic pathway gene expression profiles. Genes associated with pathways whose fluxes are upregulated have been compared with the acetate u_i cybernetic variable and genes associated with pathways whose fluxes go down have been compared with the glucose variable. Comparing the changes in gene expression with the cybernetic variables for genes related to different metabolic modules demonstrates that cybernetic variables predict the general trends in gene expression for their appropriate pathways. Explicitly, the genes go from a sustained high state to a sustained low state. Before making the comparison, it must be noted that micro-array data is inherently noisy. Because of this, many genes from these pathways were filtered out as their true behavior was hard to distinguish from the scatter in the data. The less noisy genes are plotted for various pathways in figure 4.3. For a complete picture of the data for these pathways, refer to the appendices.

Starting with secretion of acetate via the reaction of acetate kinase (ackA), the cybernetic variable for the glucose pathway was compared with dynamic transcriptional data. It is evident that both show a downward shift from a higher state to a lower one. The gene expression levels for acetate uptake via acetyl-CoA synthetase, acs, were also surveyed, however, an abundance of noise in the data made comparison difficult. However, acs is compared with less noisy data using the RT-PCR data in the next section.

Next, the cybernetic variables for glycolysis and gluconeogenesis were compared with their respective model variables. Reactions for gluconeogenesis were first compared to the acetate control variable. Most transcriptomic data for the gluconeogenic pathway were also noisy with the exception of yggF, fructose-1,6-bisphosphatase, which is somewhat less noisy. This transcript profile matches the cybernetic variable's increase. However, after reaching its highest value, the trend does not stay at an elevated state and goes down to a value roughly between the initial low state and the maximum level.

Transcripts of the glycolytic pathway show a decreasing trend as predicted by the glucose pathway's cybernetic matching law variable. For example, mRNA levels for

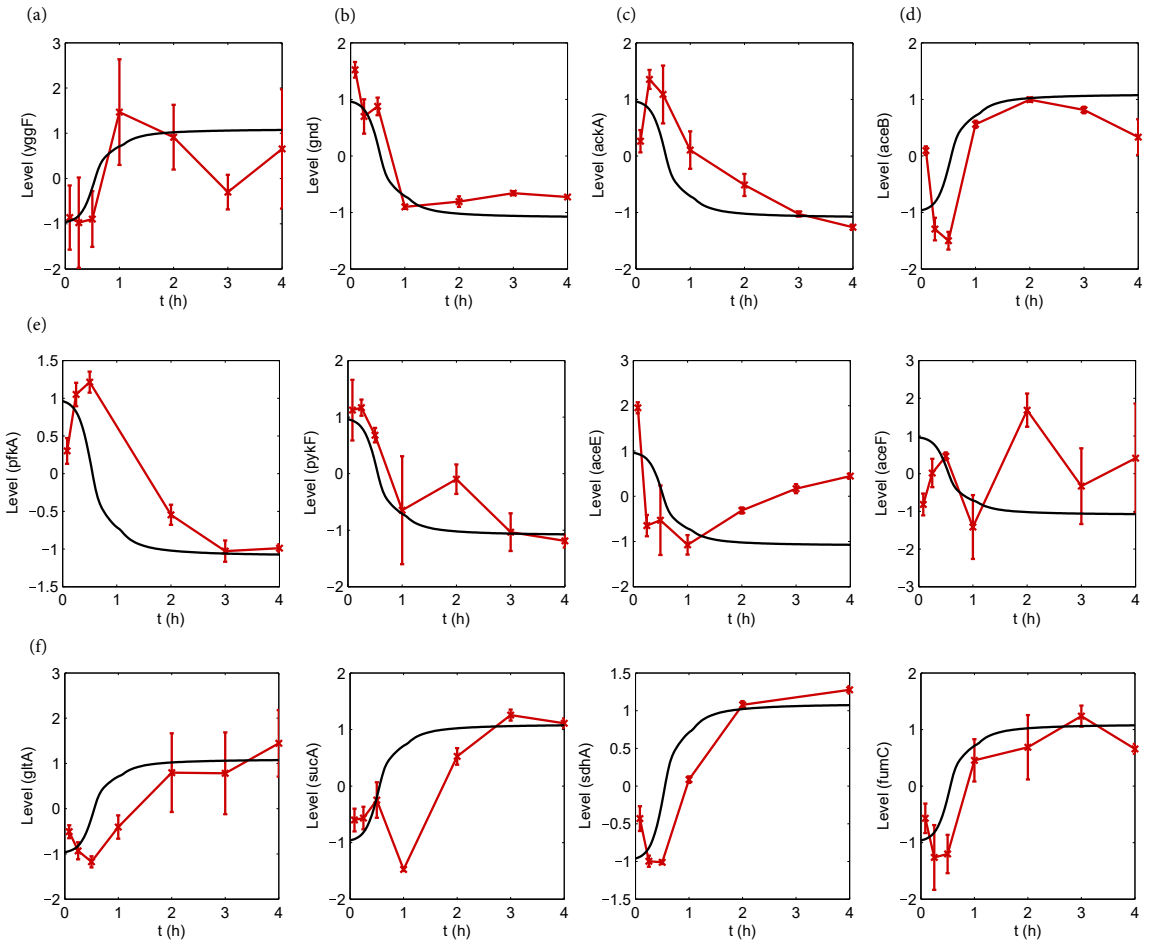

Figure 4.3.: Gene expression profile comparison with cybernetic variables for subsets of central carbon metabolism for strain MG1655's growth on glucose and acetate. Gene profiles are plotted in red with error bars and profiles are plotted in black. Subfigures with their respective pathways are labeled: (a) gluconeogenesis, (b) pentose phosphate pathway, (c) acetate secretion (d) glyoxylate pathway, (e) glycolysis, and (f) TCA cycle. Gene names are labeled on the verticle axis.

pfkA go down along with the model control variable. Also showing this behavior is pykF. Genes for the pyruvate dehydrogenase complex show varying behavior. aceE transcription levels go down as the control variable does, but rebound after 2 hours to a middle value. aceF shows significant fluctuations up and down and deviates significantly from the behavior predicted by the glucose pathway's cybernetic variable.

Finally, the TCA cycle and glyoxylate pathway were compared to the acetate control variable's behavior. Of the TCA cycle transcript data that did not have significant noise, all the genes show trends close to the model prediction. sdhA and fumC show increasing levels very similar to the selected control variable. gltA, sucA, and sucB also show general increases in expression over time. However, these expression level increases are delayed to varying degrees compared to the acetate pathway's variable. They are also characterized by an initial decrease in the first 15-30 minutes of the data series. The glyoxylate pathway also shows a general increase compared to its model prediction. Overall, the model's prediction of control of enzyme synthesis shows good conformance with what is seen at the level of gene expression.

4.3.3 Verification of Microarray Result Using RT-PCR Data from the Literature

To cross-validate the findings for the glucose-acetate diauxie previously modeled above, a second model was developed to describe a similar scenario of acetate production during the consumption of glucose followed by acetate consumption after glucose exhaustion. RT-PCR data was collected for the transcripts with the largest changes in expression for comparing steady-state growth on glucose and growth on acetate. For this analysis, this data has been partitioned into two sets. One set shows an increase in expression after glucose exhaustion corresponding to the acetate u_i variable and another represented by the glucose control variable shows a decreasing trend. A depiction of this comparison is in figure 4.4.

In this system, the data for each transcript was not analyzed according to its pathway but its general behavior overall. The genes corresponding to each control variable are listed in table II. Looking at the Pearson correlation between the vector

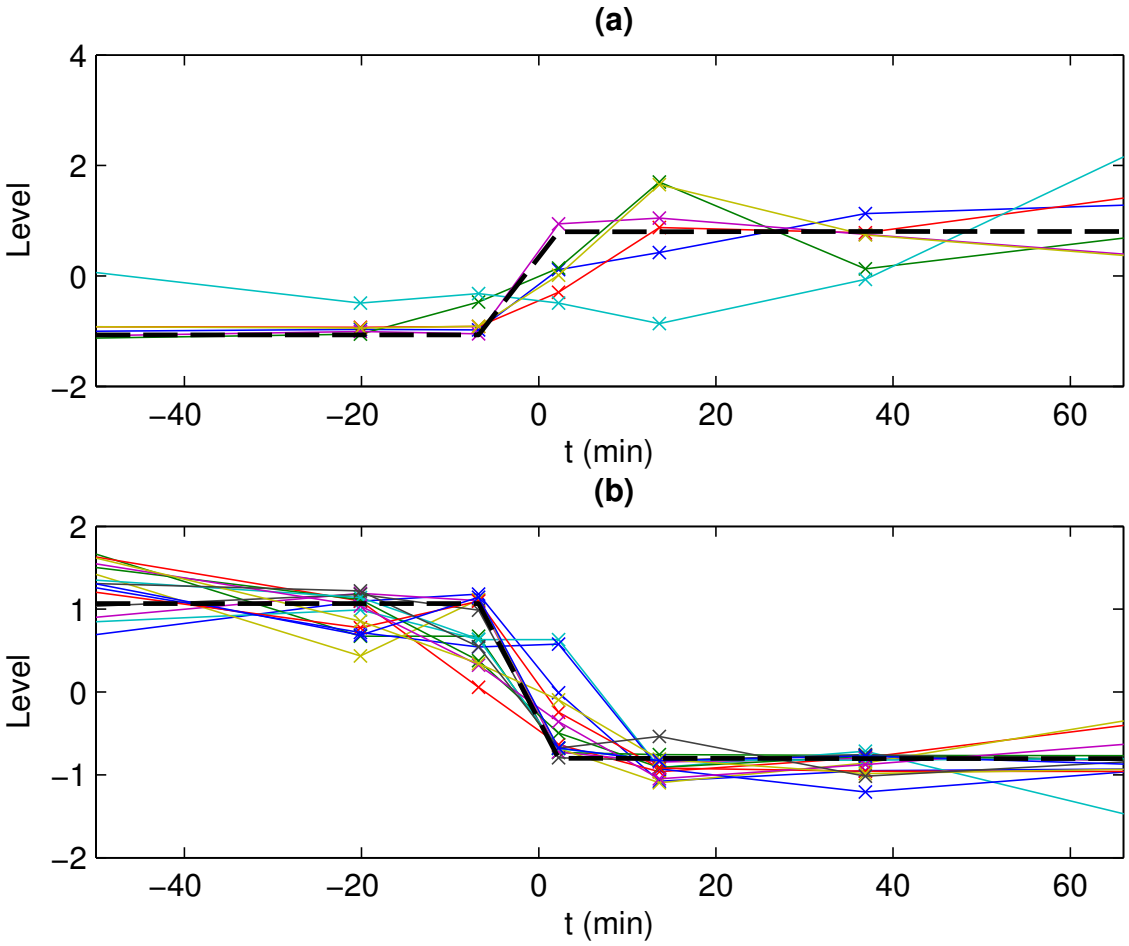

Figure 4.4.: Gene expression profile comparison with cybernetic variables for subsets of central carbon metabolism for strain MG1655's growth on glucose and acetate. Gene profiles are plotted in red with error bars and profiles are plotted in black. Subfigures with their respective pathways are labeled: (a) gluconeogenesis, (b) pentose phosphate pathway, (c) acetate secretion (d) glyoxylate pathway, (e) glycolysis, and (f) TCA cycle.

Table 4.2.: Grouping of genes with u_i variables that show similar behavior for RT-PCR data for glucose-acetate diauxie.

Acetate Group (Up)	Glucose Group (Down)	
acs	*crp*	*pfkA*
frd	*cspA*	*ppc*
glgS	*cya*	*ptsG*
ihfA	*edd*	*pykF*
pck	*infA*	*sdhD*
yfiA	*lacZ*	*sucA*
	maeA	*zwf*
	mdh	

of model u_i values at different time points and the vector representing each gene's normalized expression level, there is a strong correlation between the model prediction of enzyme synthesis control and gene expression. Overall, these correlations are 0.9214 and 0.7814 for the glucose and acetate pathway with p-values of 4.351e-044 and 1.017e-09 respectively.

After measuring the general correlation between the gene expression data and the cybernetic variables, the grouping of genes into upregulated or downregulated categories was further analyzed. Starting with uptake reactions, RT-PCR data for acs for acetate uptake and ptsG for glucose uptake show appropriate grouping with their matching law variable. The glucose variable is also grouped appropriately with genes like pfkA, and pykF from glycolysis. As expected, zwf of the pentose phosphate pathway also shows similarity to the glucose control variable. Genes strongly related to the control of growth during glucose consumption, crp and cya also show a strong similarity to the glucose control variable [71].

As for the acetate control variable, pck representing the conversion of OAA to PEP is also upregulated similar to what is expected when the cells retool their metabolism for growth on acetate. This is a key entry point of carbon into gluconeogenesis. The genes of fumarate reductase, frd, of the TCA cycle are also upregulated stepwise along with the control variable for the acetate pathway. Other genes upregulated with this control variable are glgS, ihfA and yfiA related to biofilm formation [72], DNA regulation [73] and translation inhibition [74].

There were also some unexpected relationships between the glucose variable and various components of the central carbon metabolic pathways. TCA cycle transcripts for sucA and sdhD showed a downshift. Also indicating a downshift was maeA of gluconeogenesis which convert members of the TCA cycle to gluconeogenic products. This is further analyzed in the discussion.

4.3.4 Study of Glucose-Lactose Diauxie

To verify the relationships of the enzyme synthesis cybernetic variables to gene expression in other diauxic growth systems, a glucose-lactose growth scenario for *E. coli* MG1655 was also studied. After glucose has been exhausted, genes associated with the digestion of lactose located of the *lac* operon are upregulated to facilitate growth on the new substrate. From the comparison of cybernetic u_i variables with the expression data in figure 4.5, it is clear that the lactose cybernetic variables and the regulation of genes related to the lactose pathway have a relationship. Both indicate an increase from a low to high value.

Given that lactose is a disaccharide composed of glucose and galactose, there is much overlap between the metabolic pathways for growth on lactose and growth on glucose. The main element not involved in lactose metabolism, the PTS glucose transporter complex consisting of ptsG and crr, was surveyed [75]. The behavior of this gene shows some decrease as glucose is exhausted shortly before the glucose control variable goes down. However, afterwards, it rebounds upward and shows much

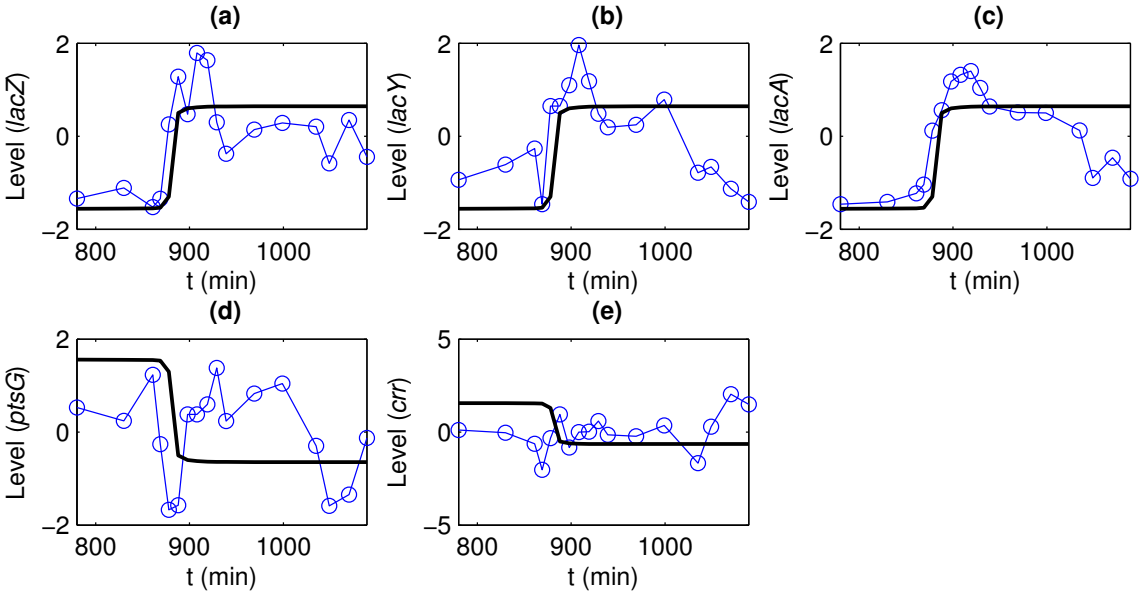

Figure 4.5.: Glucose-lactose matching law variables with gene expression profiles for relevant genes. (a)-(c) in the top row show the expression profiles of *lac* operon genes lacZ, lacY and lacA along with the lactose pathway's u_i variables. The bottom row shows the profiles for genes relevant to glucose metabolism which are the phosphotransferase glucose transporter complex of (d) ptsG and (e), regulator crr.

fluctuation up and down with decreases in expression being related to decreases in growth rate indicated on the biomass curve.

4.4 Discussion

After making comparisons between the cybernetic variable for induced enzyme synthesis, u_i, and gene expression data for three separate growth scenarios, it is evident that the two entities are related. Overall, the similarities between gene expression dynamics and the predicted control are qualitatively similar and occur on the same time scales. Moreover, when looking at correlational statistics between the least noisy data set, the RT-PCR scenario, and the cybernetic control values, a strong statistical correlation is established. For the glucose variable showing a step-down behavior, the Pearson correlation between the model and each gene related to that pathway is 0.9214 with a p-value of 4.351e-044. For the acetate pathway, it is 0.7814 with a p-value of 1.017e-09. These significant correlations indicate that there is a strong similarity between the dynamic behavior of the gene expression and the model's prediction of enzyme synthesis regulation.

It should be explicitly stated that the prediction of gene expression changes by the model are not made using any information from the gene expression level. The only input into each model's development and parameterization comes from time series data describing biomass and substrate levels. The prediction of gene expression is only made from the assumption that enzymes for substrate pathways are regulated in such a way to optimize the rate of formation of biomass. The fact that this assumption endows the model with an ability to predict changes in gene expression via the analysis of its enzyme synthesis control variable helps to validate the idea that modeling *E. coli* cells from a goal-oriented perspective is useful.

Modeling *E. coli* growth using a cybernetic approach differs substantially from other modeling methods. Instead of exhaustively enumerating the minutiae of metabolic regulatory mechanisms as is done in a host of other, kinetically driven modeling frameworks for this scenario [29, 76], cybernetic models instead model the sum of

these regulatory actions as cohesive, integrated machinery that attains some optimal behavior. By comparing the dynamics of *E. coli*'s enzyme production machinery with the cybernetic model's prediction of enzyme synthesis regulation, the wealth information related to gene expression regulation is compressed down into a succinct description related to the organism's goal. Unlike FBA which also makes assumptions regarding optimal yields [33], cybernetic models also retain considerations for the dynamic nature of optimality. Cells cannot be entirely prescient and, instead, react spontaneously to their changing environments.

The present work only has surveyed the ability of lumped cybernetic models to predict changes in gene expression. More advanced cybernetic frameworks such as the hybrid cybernetic model (HCM) [19] and the lumped hybrid cybernetic model (L-HCM) [24] may be able to generate more detailed and accurate predictions of gene expression in these scenarios. This stems from the fact that these models may incorporate multiple pathways in the form of elementary modes to describe the consumption of a single substrate. The regulation of numerous metabolic pathways could provide a higher resolution prediction of gene expression. The development of such descriptions, however, would mandate more detailed information at the level of substrates and products.

Despite an overall success of the cybernetic models in predicting gene expression changes, some discrepancies when comparing these different systems did arise. For example, the glucose-acetate microarray data for BW25113 indicated increases in expression for most all of the TCA cycle genes including frd, sucA, sdhA and mdh. In the RT-PCR data for the MG1655, the transcripts for TCA cycle genes have varying behaviors. frd shows an increase similar to the BW25113 strain data. However, sucA, sdhD, and mdh show declining expression. For these inconsistencies, it should be noted that the time periods over which the two data sets are taken varies. The MG1655 data is taken over only one hour after glucose exhaustion. On the other hand, the data analyzed in this scenario for the BW25113 strain goes for 3 hours longer than the MG1655 data. Both data sets do indicate an initial decrease. The data taken

over a longer time period demonstrate that these genes ultimately reach higher levels of expression as indicated by fluxes taken at steady states for these growth scenarios. The similarity of these gene expression profiles to the control variables is still notable. Depending on the time-frame selected for this comparison, there may be a varying but relatively small number of genes that change groupings among the different pathways. Considering the expression data from a broader angle, a significant of agreement remains between the control variables' behavior and gene expression dynamics for a majority of the genes surveyed.

The glucose-lactose system showed good agreement between the lactose pathway's cybernetic variable and the gene expression for genes dedicated to lactose metabolism. However, the glucose variable did not show a strong relationship with gene expression for genes related strictly to glucose metabolism. This is, in part, due to the fact that glucose and lactose metabolic pathways have significant overlap. Particularly, the entire pathway for glucose consumption is also used for lactose. Isolating particular chemical reactions for glucose's non-overlapping metabolic reactions is a difficult task. The main component in the glucose pathway that is not in the lactose pathway is the transporter ptsG. The expression of ptsG does not behave the same way the glucose variable behaves. It is not only expressed when glucose is present in culture, but also when lactose is being consumed. At first glance, the expression of this transporter appears to be linked more closely with growth rate and not glucose presence. When looking for explanatory information from the literature however, ptsG expression is regulated by a number of intracellular factors including mlc, crp-cAMP and Fis [77–79] and such conclusions require more consideration to develop. To model this situation more accurately, a model that dedicates a larger portion of cellular resources towards non-growth functions such as the production of ptsG could be used. Other formulations of cybernetic models have considered the idea of fractional resource investment into preadapted pathways based on other regulatory mechanisms which may be another factor related to gene expression for the glucose pathway [80].

While not all genes are exactly predicted by the trends in the u_i cybernetic variables, it is apparent that many do have a strong relationship with the model's projection of their dynamic behavior. This is especially true of genes related to the uptake of various substrates including ptsG and acs in the glucose-acetate scenario and for the lac operon of the glucose-lactose scenario. The foregoing work verifies this assumption that cells dynamically regulate different metabolic pathways towards optimality. The assumed, cybernetic mechanism of enzyme synthesis control correlates strongly with how actual biological systems modulate enzyme synthesis. This indicates that cybernetic models have the potential to predict gene expression dynamics in other systems using information at the metabolite level and an assumption of the organism's goal for metabolic regulation.

5. INVESTIGATING THE RELATIONSHIP BETWEEN CYBERNETIC VARIABLES AND TRANSCRIPTOMIC DATA: MAMMALIAN SYSTEMS

Summary

Prostaglandins (PGs) are lipid mediators that have been studied extensively due to their role in the generation of inflammation. In this study, we investigate how goal-directed models can be applied to this system to make predictions about (1) the dynamics of the conversion of arachidonic acid into downstream PG products and (2) the underlying changes in gene expression that regulate the fluxes of PG production. The behavior of eicosanoids are modeled in four separate RAW 264.7 macrophage cell treatments, a control, statin drug, KLA, and combined statin and KLA, using a cybernetic objective of maximizing the rate of production of key inflammatory mediator TNF-alpha. The model is trained on metabolite data for three of the conditions and used to predict the behavior of the fourth. The model makes predictions of induced enzyme synthesis regulation based off of the metabolic objective function and metabolite data which are compared with gene expression data. This model, for the first time, characterizes the dynamic regulation of PG synthesis and is subsequently used for applications in predicting how various cyclooxygenase inhibitors affect PG metabolism. This work embodies a novel departure from typical cybernetic modeling paradigms which typically focus on central carbon metabolism in bacterial cells. Furthermore, a generalized method for the use of different objective functions is presented for applications in mammalian cells. Overall, this work provides a model with an established ability to robustly predict perturbations to prostaglandin metabolism.

5.1 Introduction

Cybernetic models embody a framework that describes metabolic regulation from the perspective that chemical reactions inside the cell are controlled to achieve goals related to an organism's survival. For example, the regulation of metabolism can be modeled using the goal that bacterial cells tune their modulation of enzyme activity and synthesis to maximize the rate of substrate uptake [19] or the rate of growth [3]. Metabolic goals such as these have been of great utility in predicting a number of metabolic phenomena in bacterial systems including complex substrate uptake patterns [3, 5, 9], metabolic fluxes [26] and the behavior of gene knockout strains [16].

The key advantage of cybernetic descriptions of cellular regulation is that they truncate the multitude of molecular phenomena that control metabolic fluxes into an intuitive regulatory principle that guides all of these individual actions. From the cybernetic perspective, regulatory mechanisms at the molecular level are not isolated events. They are small steps in a cooperative cascade of molecular incidents that are coordinated to enhance a cell's survival. Regulatory goals, such as maximizing growth [3] or carbon uptake rate [23], provide a causality driven basis for the regulation of individual chemical events. In the absence of high resolution, dynamic data for all cellular events that modulate metabolism, cybernetic assumptions of regulation offer a significant advantage in that they are simple and can robustly predict metabolic phenomena given an appropriate objective function.

While cybernetic models have focused on bacterial systems in the past, we presently adapt this framework to model the dynamic behavior of prostaglandin formation in mammalian cell line, RAW 264.7 macrophages taken from *Mus musculus*. Prostaglandins (PGs) are a well characterized set of inflammatory lipids derived from arachidonic acid (AA) which have signaling roles. They are widely studied due to their influence in a large number of cellular events including inflammation [81] and the prevention of their formation is the intent of a large swath of cyclooxygenase (COX) inhibiting drugs [82].

Several kinetic descriptions of PG formation precede this work [83–86], but none take into account the regulatory phenomena present in PG formation. To remedy this, we develop a cybernetic description of PG formation which describes the regulation of PG synthesis as a goal-directed process related to maximizing inflammation in response to the presence of a stimulus indicative of infection.

In the development of a cybernetic model for PG formation, we seek to accomplish several outcomes. The first is to demonstrate that cybernetic models are a robust description of metabolite formation and can be used to predict perturbations to metabolism via various effectors including drugs. Having a more reliable description of PG formation is useful in that it can provide more predictive description of the action of inhibitory drugs. Second, provide proof that cybernetic models can be used to predict gene expression phenomena in complex mammalian systems which has been shown previously in diauxic growth scenarios in chapter 4. Showing an ability to predict gene expression allows for validation of the PG model's objective function and more generally, cybernetic models. Lastly, it is the goal of this work to offer proof of concept that return on investment can be broadened to describe objective functions in complex multi-cellular systems and have applications in predicting the response of metabolic networks to drugs. Prior cybernetic formulations have used objective functions based off of arguments related to survival in competitive environments to model the behavior of individual bacterial cells. In describing the behavior of cooperating cells in multi-cellular environments, refinement of cybernetic objectives is needed.

In achieving the above, a new paradigm of cybernetic modelling is provided, one that endows its users with the ability to make predictions in pharmaceutically relevant systems. This opens up the potential for more robust modeling of the dynamics of human metabolism using cybernetic models with a more refined interpretation of objective functions. Also, the validation of cybernetic control variables using gene expression data demonstrates the utility of this framework in providing mechanistic explanations for the behavior of bioinformatic data.

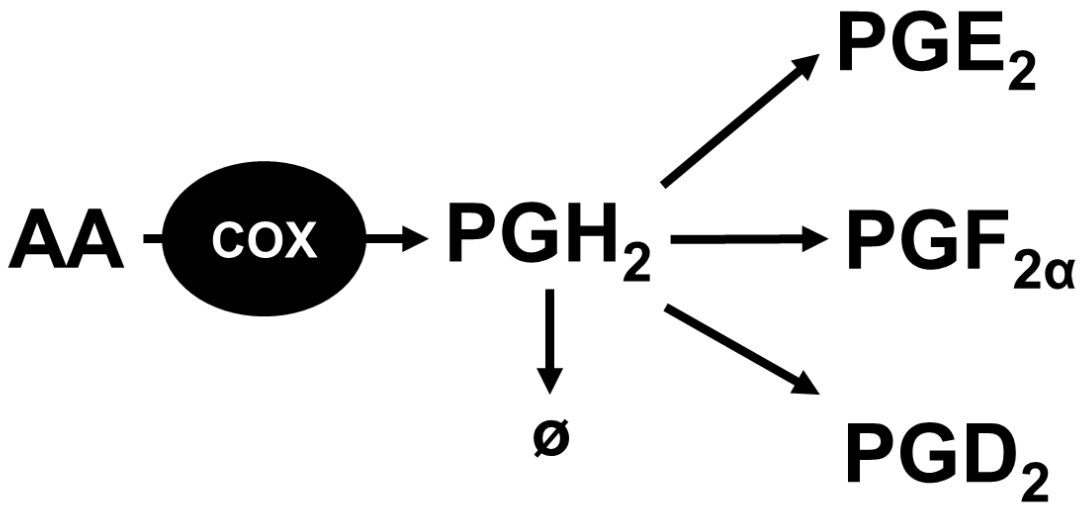

Figure 5.1.: Schematic of prostaglandin network modeled.

5.2 Materials and Methods

To describe the time-dependent formation of prostaglandins, a cybernetic model is generated. This description approximates the conversion of AA into intermediate product, prostaglandin H_2 (PGH_2) and its subsequent conversion into downstream prostaglandin E2 (PGE_2), prostaglandin F_{2a} (PGF_{2a}), and prostaglandin D_2 (PGD_2). To see the structure of the reaction network modeled, refer to the network schematic in figure 5.1. The behavior of this network is modeled in four separate conditions, a control, a treatment with compactin, Kdo_2-Lipid A (KLA), and a combined treatment of compactin and KLA. Compactin is a type of statin that reduces cholesterol formation by blocking HMG-reductase CoA and also influences prostaglandin formation. KLA, is an LPS analogue that signals the presence of infection to the macrophage cells and stimulates the formation of prostaglandins by binding to a receptor on the macrophage called Toll-Like Receptor 4 (TLR4) [87, 88]. The data for all of these conditions was taken from [89].

In this simple network of PG formation, the main focus is on how PGH_2 is converted into downstream PGs because the regulation of this branch point involving the synthesis of three separate products represents a central decision point in this small metabolic system. In using cybernetic arguments to model PG formation, we are assuming that these products are formed in varying amounts related to their ability to help the cell achieve its metabolic objective. The production of PGs that have a stronger relationship with the goal of the system will be upregulated while the pathways for those PGs which have a lesser relationship with the objective function will be downregulated. Using goal-directed arguments such as these, gives the model generated here a more holistic perspective of cellular regulation as its regulation pertains to the physiological function of the metabolic network itself. The network generates metabolites in order to accomplish some ultimate function which is embedded into the model using cybernetic regulation. To understand cybernetic regulation more deeply, let us first focus on the model's interpretation of metabolic fluxes.

5.2.1 Reaction Kinetics

The structure of the kinetics for this reaction network are simplified into two segments. The first describes the conversion of AA into PGH_2 using reaction kinetics that do not involve cybernetic regulation. This is a reasonable simplification to make in that there is a single pathway that leads from one reactant to one single product. Despite its inability to be measured experimentally, PGH2 is preserved in this model as it is a central intermediate for guiding the formation of PGs downstream and there is a necessary delay between the consumption of AA and the formation of downstream PG products. The kinetics of this reaction are modeled as three separate mechanisms for PG formation including a basal rate of synthesis and stimulated generation due to compactin, $COMP$ and KLA, KLA.

$$r_{AA\rightarrow PGH_2} = \frac{k^{\max}_{H_2,base} AA}{K_{H_2} + AA} + \frac{k^{\max}_{H_2,comp} AA}{K_{H_2} + AA}[COMP] + \frac{k^{\max}_{H_2,kla} AA}{K_{H_2} + AA}[KLA] \tag{5.1}$$

which simplifies to

$$r_{AA\rightarrow PGH_2} = \frac{k^{\max}_{H_2,base} AA}{K_{H_2} + AA}(1 + k'_{H_2,comp}[COMP] + k'_{H_2,kla}[KLA]) \tag{5.2}$$

This additive mechanism is a simplification of other possible kinetic formulations that involve cofactors which increase activity. For further simplicity, these kinetics involve the same Michaelis-Menten parameter for all three, additive mechanisms for prostaglandin formation. Testing other kinetic formulations of this system did not endow the model with an enhanced ability to fit or predict the changes in metabolic data in a substantial way.

The other segment in this model employs cybernetic regulation as regulation between the different metabolic options is the central interest in this work. In cybernetic models, there are two descriptions of the reaction kinetics. One is the raw, enzyme-dependent rate of reaction which will henceforth be termed the kinetic rate of reaction. This kinetic rate includes an enzyme quantity e_i which represents the relative amount of enzyme devoted to the conversion of PGH_2 to a PG product.

$$r^{kin}_{PGH_2\rightarrow PG_i} = e_i \frac{k^{\max}_{i,base} PGH_2}{K_i + PGH_2 + K_{inh} PGH_2^2}(1 + k_{i,comp}[COMP]) \tag{5.3}$$

This kinetic rate employs substrate inhibition constant K_{inh} in its formulation as the metabolite data indicates that when compactin stimulation is present, there is a decreased level of substrate but increased level product formation. Stimulation by KLA for these downstream pathways was omitted because KLA stimulation upstream provided a sufficient description of the metabolite data for the model. The second reaction rate in this model is the activity regulated reaction rate which is modulated by the cybernetic control variable v_i

$$r^{reg}_{PGH_2\rightarrow PG_i} = v_i e_i \frac{k^{\max}_{i,base} PGH_2}{K_i + PGH_2 + K_{inh} PGH_2^2}(1 + k_{i,comp}[COMP]) \tag{5.4}$$

More detailed information on the formulation of cybernetic control variables and the regulation scheme used in this model can be found in the subsequent section on cybernetic variables.

5.2.2 Modeling Metabolite Changes

In this model, the changes in metabolites are modeled by several different mechanisms. The upstream production and downstream consumption of the AA substrate in this network is not modeled due to the fact that the upstream generation of AA comes from several pools of upstream resources and is not the focus of this work. To describe the time dependent changes in AA, a simple time-dependent function is fit to the AA data using cubic spline interpolation. The generation of downstream metabolites PGH_2, PGE_2, PGF_{2a}, and PGD_2 from the consumption of AA are described using differential equations with generation and decay mechanisms. For all PGs in the model, first order decay processes are used with decay constants γ_i. These decay values represent the amount of the metabolite pool that are converted to other metabolites which are not modeled due to lack of data for alternate reactions such as thromboxane A and PGI_2. The changes in PGH_2 are modeled as

$$\frac{dPGH_2}{dt} = r_{AA\rightarrow PGH_2} - \sum_{i=1}^{3} r_{PGH_2\rightarrow PG_i} - \gamma_{PGH_2} PGH_2 \tag{5.5}$$

Similarly, the changes in downstream PGs are modeled with

$$\frac{dPG_i}{dt} = r_{PGH_2\rightarrow PG_i} - \gamma_i PG_i \tag{5.6}$$

Note that all metabolite concentrations described are normalized by the amount of DNA taken from the culture and are in the units of pmol/ug DNA. Because of this, the influence on cell concentration is ignored in the kinetics of this model this model. In addition to the changes in metabolites, the relative changes in enzyme level, e_i, for each pathway are also modeled with first order ODEs.

$$\frac{de_i}{dt} = \alpha_i + u_i \frac{k_{E,i} s_i}{K'_i + s_i} - \beta_i e_i \tag{5.7}$$

Three phenomena, constitutive formation, induced formation and degradation, are included in the description of changes in enzymes. The cybernetic variable u_i represents the regulation of induced enzyme formation which will be given much attention in later discussion.

Lastly, to explain the changes in compactin and KLA, factors that stimulate the formation of PGs, the treatments to the culture are modeled in a similar fashion to [83] with a piecewise function. This piecewise function ramps up to a maximal value of 1 at 0.5 hours and decreases exponentially afterwards.

$$KLA = \begin{cases} 2t \text{ if } x \leq 0.5 \\ e^{-k_d t} \text{ if } x > 0.5 \end{cases} \tag{5.8}$$

The main departure of this piecewise function from previous work is that the second term includes exponential decay instead of a linear function to describe the desensitization of the cells to the stimulus which is modeled according to the dynamics measured in [90].

5.2.3 Cybernetic Regulation for Novel Objective Functions

The distinguishing element of cybernetic models is their approximation of the cellular control of metabolic reactions through the modulation of enzyme quantity e_i and changes in activity for reaction enzymes. The fundamental assumption of cybernetic models is that cells control metabolism towards goals related to their survival and propagation. Traditional cybernetic models with goals such as maximizing carbon uptake rate or the rate of growth are intuitively related to the natural fitness of organisms. In the present work, we focus on the PG metabolic system which is significantly smaller than central carbon metabolism. Despite differences in scale, the concepts of cybernetic modeling remain applicable in that the efficient modulation of subnetworks in the overall metabolic network is important to the overall survival and function of an organism. PGs may not generate growth themselves, but the efficient management of their production by the cell is an important component of the cell's function and its host organism's overall survival.

PGs have an array of functions in multicellular organisms. After being generated by the action of either COX-1 or COX-2 on AA, they are released into the cellular medium and subsequently bind to PG specific G-protein-coupled receptors on the cell surface [81, 91]. These receptors, such as EP_1 through EP_4, FP, and DP_1 and

DP_2 activate the formation of second messengers such as Ca^{2+} and cAMP. These second messengers subsequently exert a number of varied effects on the cell. Generally speaking, PGs are well-characterized for their roles in the formation and resolution of inflammation. In this particular exercise, we will focus in on PGs' role in the generation of inflammation for the selection of the model's objective function.

While there is no single entity that represents the totality of inflammation alone, the cytokine tumor necrosis factor alpha (TNF-alpha) is well-known for its role in the generation of systematic inflammation and is a product of the response to LPS [92]. Because of TNF-alpha's well known role as a central mediator of inflammation [93], it is our hypothesis that the metabolism of PGs is regulated to maximize inflammation which characterized by the amount of TNF-alpha generated by the system. To quantify the relationship between each PG and TNF-alpha, a simple, linear model is used that relates the time-dependent changes in TNF-alpha level with some weighted combination of the time-dependent changes in PGs as

$$T\dot{NF}\alpha = \sum_i w_i P\dot{G}_i \tag{5.9}$$

where w_i is some weight that is fit using least-squares regression. What this objective function means is that of the three pathways modeled, there is a varying amount of inflammation that results from the generation of each PG_i. It is conceivable that other mediators of inflammation may be more relevant as objective functions such as IL-6. However, to justify our approach from the basis of biomolecular signaling pathways, TNF-alpha is induced by NF-κB [94], a factor that is stimulated by Ca^{2+} [95], a factor which is stimulated by various PG receptors [96].

In the cybernetic framework, the production of enzymes for pathways associated with the objective function will be upregulated while enzymes for pathways that are suboptimal in maximizing the objective function will be downregulated. This is calculated using cybernetic variables which compare returns on investment (ROI) for the various pathways

$$u_i = \frac{p_i}{\sum_j p_j} \tag{5.10}$$

In this particular system, the ROI for each pathway is the assumed to be the amount of TNF-alpha that each unregulated pathway can yield at each instant in time which is

$$p_i = w_i r^{kin}_{PGH_2 \rightarrow PG_i} \tag{5.11}$$

assuming no delay in between the formation of PGs and the downstream synthesis TNF-alpha. This assumption of no delay is a remedy for the fact that there is little quantitation of the actual delays between these two systems and the time scales modeled are much larger than the relative magnitude of time over which this delay would occur. The u_i variable modulates the induced enzyme synthesis rate in equation 5.7. In essence, this cybernetic variable states that the fraction of the limited pool of resources devoted to the synthesis of PG generating enzymes is proportional to the ROI generated by making enzymes for that pathway.

The other form of regulation that is used in this model is for the regulation of enzyme activity of different enzymes which typically occurs through allosteric mechanisms. This proportional variable is formulated as

$$v_i = \frac{p_i}{\max\limits_j (p_j)} \tag{5.12}$$

with the same returns on investment. To put the meaning of the proportional variable into words, the pathway with the highest ROI is fully expressed while suboptimal pathways have their activity proportionally downregulated as their full expression consumes resources in a less-than-optimal fashion.

5.2.4 Model Parameterization

To first develop the model, the objective function needs to be established by connecting the PGs in the model with TNF-alpha as in equation 5.9. To calculate the linear weights, a linear least squares fitting was performed using numerically approximated derivatives for TNF-alpha and PG data from the three of the four conditions with a constraint that the linear weights must be positive. This constraint is necessary in that cybernetic regulation mechanisms cannot reverse the flow of

Table 5.1.: Parameter values for PG model

Parameter Name	Value	Parameter Name	Value
w_{PGE_2}	0.23	γ_{PGE_2}	1.77
$w_{PGF_{2a}}$	0.20	$\gamma_{PGF_{2a}}$	5.10
w_{PGD_2}	0.57	γ_{PGD_2}	1.41
k_{PGE_2}	155	$k_{PGH_2,base}$	5.33
$k_{PGF_{2a}}$	121	$k_{H_2,LPS}$	208
k_{PGD_2}	165	$k_{H_2,comp}$	0.0037
K_{PGE_2}	2.14	$k_{H_2,base}$	1.7527
$K_{PGF_{2a}}$	1.62	K_{inh}	1.65
K_{PGD_2}	5.33	α	0.004
K_{H_2}	1.74	β	0.05
γ_{PGH_2}	34.9	k_E	0.343

reactions or the induced enzyme term cannot become negative. To find the weights calculated by the model, refer to table 5.1.

The model was parameterized using data from three of the four conditions, the control, the compactin treatment and the KLA treatment. Data was available for the metabolite concentrations of AA, PGE_2, PGF_{2a}, and PGD_2 over a 24 hour time window. To fit the model to the data, a normalized least squares objective function was used to ensure that the varying magnitude of each PG's level did not skew the parameters towards the sole fit of PGs with higher magnitudes. The time intervals for the data were taken at 0, 0.5, 1, 2, 4, 8, 12, and 24 hours. To reduce the bias of the fit towards the beginning of the series where there is a higher density of data

points, each time point's data was normalized by the square root of its time value. The overall objective function for fitting the data was

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{8}\left(\sqrt{t_k}\left(\frac{y_{\text{exp},ijk}-y_{\text{mod},ijk}}{y_{\text{exp,max},ij}}\right)\right)^2 \tag{5.13}$$

where i is the experimental condition, j is the downstream PG identifier and k is a particular data point for the series. The ODEs in the model were solved using the routine ode15s for stiff systems in MATLAB (2015, Natick, MA) and the objective function was minimized using an optimization procedure that started with a genetic algorithm seeded with random initial parameter values. The genetic algorithm result was then refined using a constrained gradient search method (fmincon). After running the optimization procedure for 2000 randomly generated starting points, the final parameter values were selected as the ones that best minimized the objective function. The parameter values can be referenced in table 5.1.

5.3 Quantifying Model Accuracy

Cybernetic variables describe the regulation of enzyme synthesis. An analogue to the cybernetic variables is gene expression data which represent the concentrations of mRNA which control the formation of enzymes. While u_i variables and gene expression are not directly substitutable, the changes in cybernetic variables should ideally mimic the behavior of mRNA data. These quantities are hard to compare in an absolute sense so the variables are scaled in the following manner consistent with the scaling in chapter 4:

$$u'_i(t) = \frac{u'_i(t) - \mu(u_i)}{\sigma(u_i)}, \tag{5.14}$$

where $\mu(u_i)$ represents the mean of the time-series' values and $\sigma(u_i)$ is the standard deviation [69]. The gene expression variables are scaled in the same way. By scaling both variables in this fashion, it is easy to compare the dynamic behaviors qualitatively and through scale-invariant metrics like Pearson's product-moment correlation coefficient. Demonstrating a strong correlation between changes in cybernetic vari-

ables for enzyme formation and the changes in gene expression helps to validate the goal-directed control mechanism in the model.

5.4 Results

5.4.1 Prediction of Metabolites

After fitting parameters to three of the four conditions (i.e. the control, compactin treatment, and KLA treatment conditions), the model provided the following fits which are shown in figure 5.2. The substrate curves for AA show a close representation of the different conditions. Also, it is evident that the model correctly explains the evolution of the metabolite concentrations for the different conditions involved in the fit. The control and compactin treatment scenarios both show a relatively low rate of prostaglandin formation. The KLA treatment shows a good agreement with all prostaglandin products generated.

The kinetics of the model were cross-validated using the fourth condition, the combined compactin and KLA treatment which is shown in the blue line in figure 5.2. While the fit of the cross-validation case is not perfectly accurate, it is evident that the model predicts the qualitative behavior of the validation condition. The substrate concentration, AA, goes down, but the levels of all products go up. The predictions of the formation of PGD_2 are especially accurate. Model predictions of the behavior of PGE_2 and PGF_{2a} are under the observed data, however, there is, nonetheless, an increase in formation of these products going from the KLA only condition to the combined KLA and compaction treatment. Because of the model's ability to capture this unusual behavior without being directly informed of it, it is reasonable to conclude that the model is predictive in describing the behavior of prostaglandin metabolites.

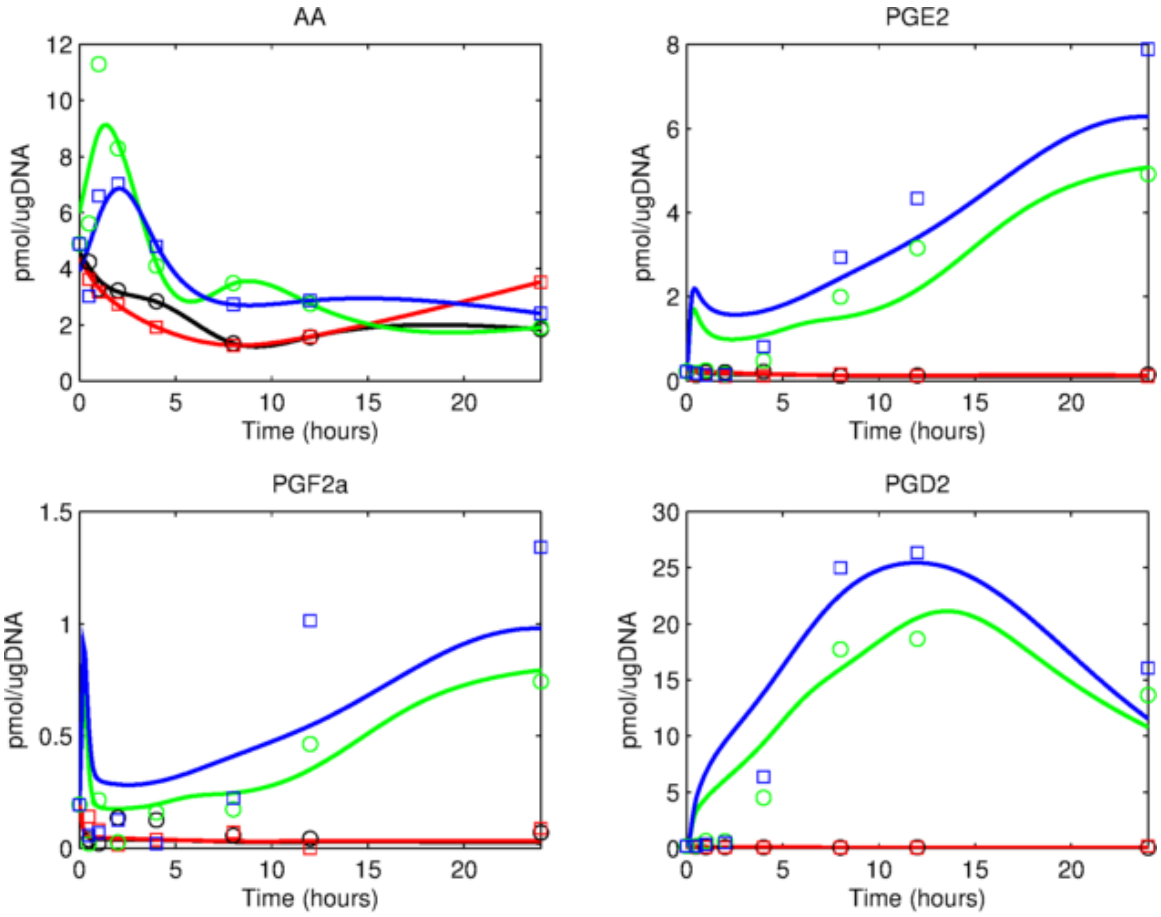

Figure 5.2.: Time evolution of metabolite concentrations (pmol/ugDNA) for prostaglandin system. Each condition is distinguished by color with the control case in black, compactin treatment in red, KLA treatment in green and combined KLA and compactin treatment in blue. Data points for the each of these conditions are in the same color.

Table 5.2.: Genes that catalyze different pathways in the model.

Entrez ID	Pathway	Gene Symbol	Name
64292	$PGH_2 \rightarrow PGE_2$	Ptges	prostaglandin E synthase
96979	$PGH_2 \rightarrow PGE_2$	Ptges2	prostaglandin E synthase 2
109857	$PGH_2 \rightarrow PGF_{2a}$	Cbr3	carbonyl reductase 3
66469	$PGH_2 \rightarrow PGF_{2a}$	Fam213b	family with sequence similarity 213, member B
19215	$PGH_2 \rightarrow PGD_2$	Ptgds	prostaglandin D2 synthase

5.4.2 Prediction of Gene Expression Changes

In order to validate the cybernetic control mechanism that drives the modulation of reaction rates in the model, scaled gene expression data, representative of the control of enzyme synthesis, were compared to scaled versions of the cybernetic variables that specify the control of enzyme synthesis. The qualitative trends among both the gene expression data and the cybernetic variables should be similar. In other words, if the gene expression for one of the pathway is increasing over a certain time period, the cybernetic variable for enzyme synthesis control should also be increasing.

Before this comparison could be made, it was necessary to identify which genes were related to the different pathways. These genes were isolated using the KEGG database [97, 98]. The genes that produce enzymes that catalyze the pathways of focus are highlighted in Table II. For two of the three pathways modeled there are multiple genes that are associated with the catalysis of a pathways in the network. This is addressed by comparing the behavior of the multiple genes with the individual cybernetic variables for the given pathways.

It is evident in Figure 5.3 that the cybernetic profiles in red match the behavior of the genes denoted by dashed and solid black lines. These comparisons are made

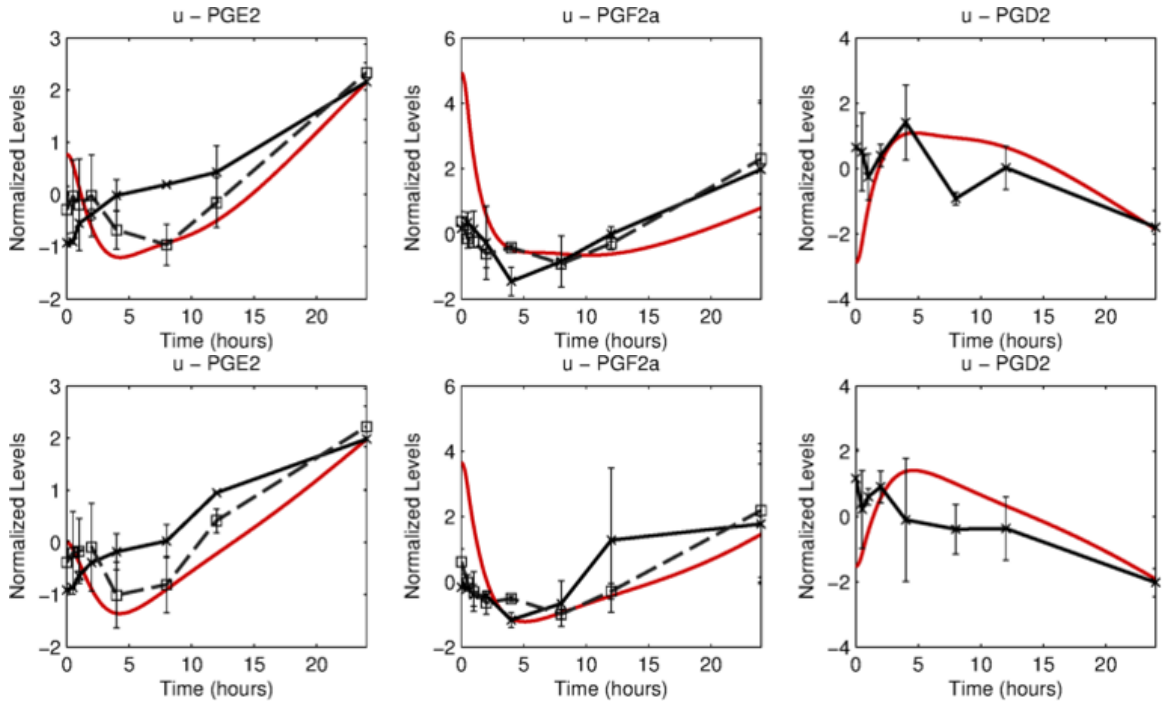

Figure 5.3.: Behavior of the scaled cybernetic variables (red) compared to scaled gene expression values (black solid and dashed lines) for PGE_2, PGF_{2a} and PGD_2 pathways in (top) KLA and (bottom) combined KLA and compactin treatments. Genes for PGE_2 are Ptges (solid) and Ptges2 (dashed). PGF_{2a} genes are Cbr3 (solid) and Fam213b (dashed). The PGD_2 gene in black is Ptgds.

for both the KLA and combined KLA and compactin treatments. They were not made for the control or compactin treatments due the to the fact that the gene expression data was very noisy. The cybernetic variable for the PGE_2 producing pathways shows a behavior where the gene expression values start at some level, go down to a minimum value and then go up. This behavior is very similar to the behavior of the prostaglandin E synthase 2 which also goes down to some minimum around 8 hours and then increases in value. The genes and cybernetic variables for the pathway that generates PGF_{2a} also show similar behavior to each where the value starts at a high value, decreases, and then goes up. The pathway for the PGD_2 pathway has a cybernetic variable that shows the value starting at some low value, increasing to a maximum around 4 hours and decreasing down. The gene expression for prostaglandin D_2 synthase also shows similar changes with the long term profile showing a downward decrease.

Towards quantitative understanding of the similarity between the gene expression profiles and the cybernetic variables, Pearson's product-moment correlation coefficients were calculated for the two scenarios without noisy gene expression, the KLA and KLA Compactin treatments. Each scaled gene expression series was compared with its respective cybernetic variable. The KLA treatment case had a statistically significant correlation coefficient (ρ=0.35 / p-value=5.2e-5). The combined KLA and compactin treatment had an even more significant statistical correlation (ρ=0.43 / p-value=5.3e-6). The significance of the correlation between the cybernetic variables and data for gene expression, a mechanism that actually controls the rates of reactions, demonstrates the consistency of the regulatory description in the model with the reality of metabolic control.

5.4.3 Predictions of Drug-Metabolism Interactions in PG Network

Given that the inhibition prostaglandin formation is the aim of a wide variety of drugs called nonsteroidal anti-inflammatory drugs (NSAIDs), the present model is capable of making predictions of how inhibiting the upstream conversion of AA to

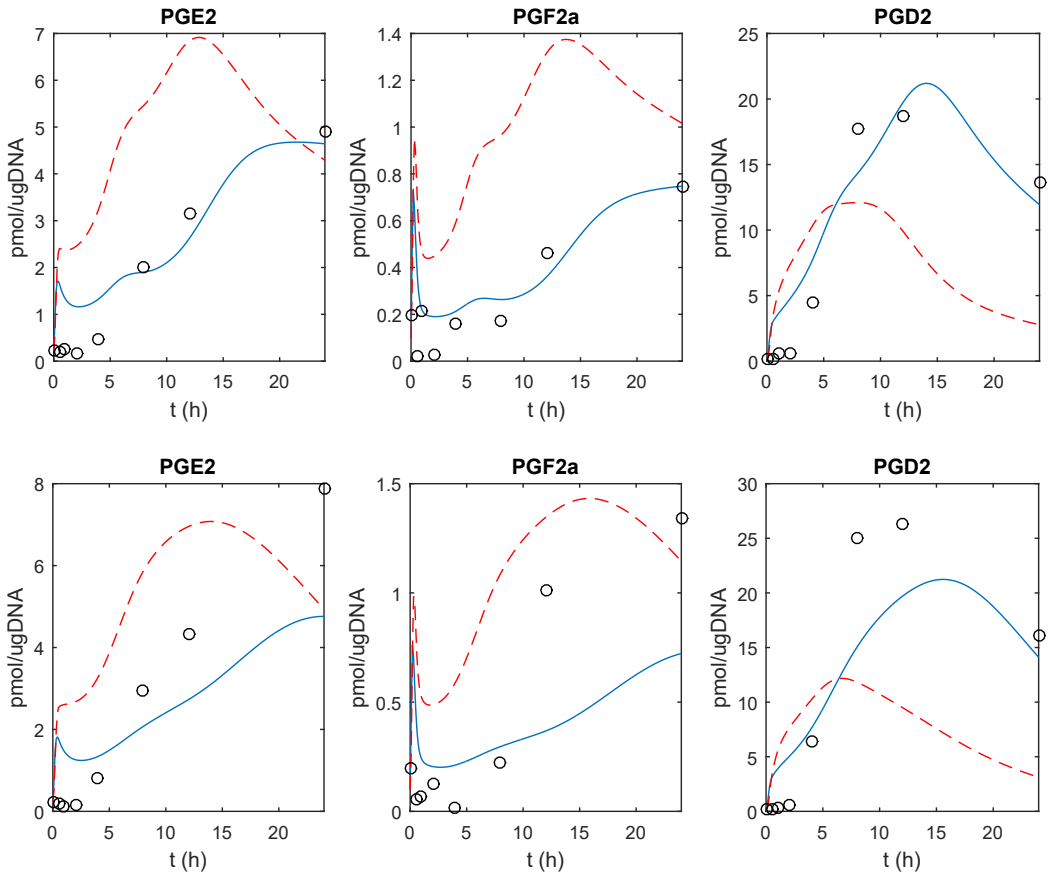

Figure 5.4.: Production of prostaglandins for KLA (top) and combined KLA and compactin (bottom) treatments. Data points for untreated conditions are signified by circles. The blue line represents the uninhibited network. The red dashed (- -) lines represent the treatment with PGH_2 inhibiting drugs.

PGH_2 affects downstream PG concentrations. The behavior of PG formation can be extrapolated through an understanding of how much cyclooxygenase is inhibited by various NSAIDs which can come from many sources [99,100]. We can use these values to characterize the effects of different drugs on the system.

Figure 5.4 shows the behavior of how 50% inhibition of cyclooxygenase affects the formation of PGs relative to the uninhibited case in red. The model predicts that there will be a significant decrease in the amount of PGD2 formed and increases in

the amounts of PGE_2 and PGF_{2a} in both the KLA and KLA+compactin cases. The behavior predicted is a function of the nonlinear kinetics used to model changes in PG concentration. This result has yet to be experimentally verified.

5.5 Discussion

The foregoing model of PGs has been validated by two levels of biological data that depict the synthesis of PGs. The kinetics that describe the rates of reaction are able to predict the generation of PGs for conditions upon which the model has not been trained. Furthermore, the control variables for enzyme synthesis show strong correlation with gene expression data. Given the fact that this model shows consistency in the prediction of different biological phenomena, it is feasible that it should provide for robust predictions of perturbations to the PG system of study.

The objective function in the model, maximizing the rate of TNF-alpha formation, is a central postulate of the model. It is feasible that other objective functions could also be descriptive of the control of PG formation. While TNF-alpha is well characterized as signaling molecule generated in the macrophage response to KLA binding to TLR4, other inflammatory cytokines such as interleukins (ILs) like IL-1, IL-6 and IL-12 could also be the goal of the system [101].

Control goals related to other functions of PGs besides inflammation could also be of interest; however, given that the response of macrophages to KLA is an inflammatory one, the objective function of TNF-alpha has meaning in the context of the system and conditions studied. To generate the cybernetic model, it is necessary to first make an assumption of what the control goal is. This assumption can then be tested on the basis of whether the model is capable of making predictions of data beyond what the model is trained on. The fact that the model, with the TNF-alpha objective function, is able to make predictions of the KLA and compactin treatment as well as of gene expression trends helps to validate the control assumption central to the model. This control goal gives deeper insight into the nature of prostaglandin generation in response to infectious stimuli in that it describes an overarching prin-

ciple of organization to metabolic regulation; the cells are tuning the generation of various prostaglandins to maximize the rate of TNF-alpha formation.

The implementation of the TNF-alpha objective function as a linear combination of PG product formation rates is made as means of convenience; the mechanism connecting PGs to TNF-alpha production is not fully characterized in the format of dynamic data. In the event that such data were available, a model connecting the upstream formation of PGs with TNF-alpha could be generated using a hybrid cybernetic model [19].

While the present work focuses on dynamics of a smaller metabolic network, the success of the model in predicting various metabolites serves as a proof-of-concept for cybernetic modeling to describe the segments of complex mammalian metabolic behavior. This shows that cybernetic models could be developed to describe the behavior of larger networks in mammalian cells as they have been generated for complex bacterial metabolic networks [26, 102]. Multicellular metabolism has already been probed using static, yield-based objective functions [103–105], but the determination of dynamic objective functions would provide for a more complete picture of the multi-cellular metabolic dynamics in more complex organisms.

5.6 Conclusions

This work, for the first time, developed the idea that cybernetic metabolic objectives can be used to describe the regulation of signaling systems mammalian metabolism. It yielded a model describing PG synthesis that is capable of predicting both metabolites and the relative changes in gene expression. This model was then used to provide robust predictions of how drugs that inhibit PGH_2 formation alter the downstream generation of PGs which also marks the first time that cybernetic models could be explicitly used for pursuits in translational research.

6. MOTIVATING THE NEED FOR A METHOD TO ESTABLISH THE UNIQUENESS OF OBJECTIVE FUNCTIONS

6.1 Introduction

In developing the model which describes the formation of Prostaglandins in the prior chapter, the regulation of the metabolic reactions relied on a postulate that the metabolic goal of the system was to maximize the rate of formation of TNF-alpha. The model was subsequently validated using data for gene expression and metabolites. Despite these levels of validation, TNF-alpha rate maximization, as the only feasible objective function for the PG system, is inconclusive.

6.2 Testing of Arbitrary Objective Functions

When analyzing the uniqueness of weights for the objective of maximizing TNF-alpha, there is an intriguing result. To reiterate the approach to modeling the eicosanoids, the basics of the implementation of the objective function will be covered.

The network for the model is quite simple as shown in figure 6.1 with the starting lipid AA being converted into downstream product PGE_2, PGF_{2a}, and PGD_2 in a simple two step reaction process that is linked with an objective function downstream for the production of inflammatory cytokine TNF-alpha.

To connect the downstream products of this metabolic network with the objective of TNF-alpha, a linear model is employed where the change in TNF-alpha is some linear combination of the time-dependent changes in PGs:

$$\dot{TNF\alpha} = \sum_i w_i \dot{PG_i} \tag{6.1}$$

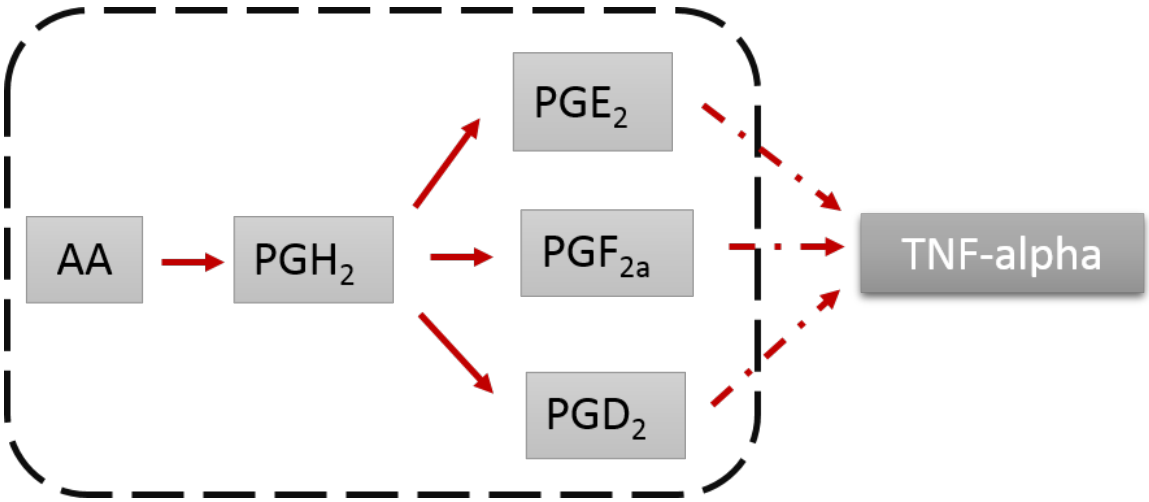

Figure 6.1.: Schematic of PG network with goal of TNF-alpha production.

These weights then are used to bias the returns on investment in the cybernetic control variable u_i

$$u_i = \frac{p_i}{\sum_j p_j} \tag{6.2}$$

which is composed of the ROIs:

$$p_i = w_i r^{kin}_{PGH_2 \rightarrow PG_i}. \tag{6.3}$$

where each p_i represents the unregulated amount of TNF-alpha that can be generated downstream through signaling processes by the full expression of each pathway. As stated in Chapter 5, this assumes that the time delay in the signaling mechanisms between prostaglandins and TNF-alpha is negligible for the time scales over which the metabolites are modeled.

Depending on the objective function that is selected, the weights w_i for each prostaglandin's production will be different. To test the uniqueness of the TNF-alpha weights, 24 sets of weights were generated as arbitrary objective functions using Latin hypercube sampling (LHS). Kinetic parameters were then fit to the data using each weight set and the resulting gene expression trends were surveyed.

The results are in figure 6.2. The u_i data is scaled so that it the whole set aligns regardless of initial values of u_i. The surprising result is that the average trends of all the normalized weight sets that were tested show changes in u_i variables that are true to the gene expression pattern. One can see in figure 6.2 that the behavior of all weight sets is more or less the same (especially for PGF_{2a} and PGD_2). This could imply that judicious selection of the objective function is not necessary because the cybernetic model can robustly model gene expression trends no matter what weights are applied as an objective function.

6.3 Estimation of Cybernetic Parameters from Data: An Inverse Problem

The goal of this segment is to further reflect upon the nature and uniqueness of objective functions in cybernetic approximations of metabolic changes. This deliberation is in response to the discovery that there is no unique set of return on investment

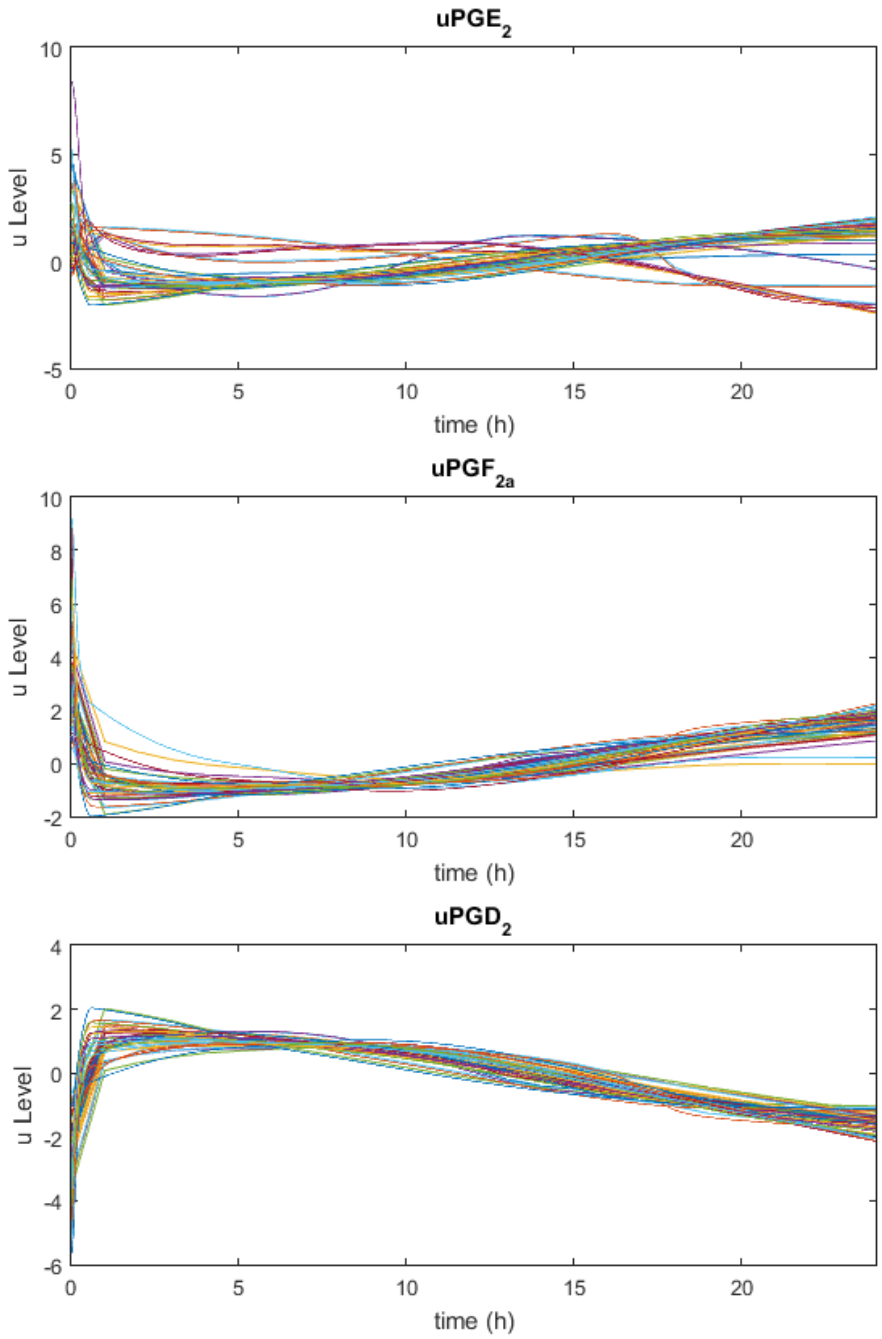

Figure 6.2.: Behavior of cybernetic matching variables for the production of prostaglandin products PGE_2, PGF_{2a}, and PGD_2 for various objective function weightings. Comparison of mean behavior of profiles compares favorably with gene profiles in figure 5.3 (a).

weightings that provide a singular description of the observed data for regulation. The central issue is that the unregulated kinetics in the model, determined by the fitting of parameters to data, tend to yield the same predictions of gene expression trends despite variations in the unregulated kinetics themselves. In other words, the inverse problem of estimating parameters from the data estimates the same trends in regulation despite different control goals.

6.3.1 Background: Objective functions

Given an inferred objective, J , and some time-series representation of this objective, one can determine a relationship between the product or intermediate metabolites of the network, m_i, using linear least-squares fitting on the time-dependent changes in J and m_i.

$$\dot{J} = \sum_i w_i \dot{m}_i \tag{6.4}$$

These weights, the manifestation of the hypothesized relationship between the objective and the metabolic system of interest can then be used to infer the regulation mechanisms that govern this system through their participation in each pathway's calculated return on investment, p_i where

$$p_i = w_i r_i^{kin} \tag{6.5}$$

This expression for return on investment then goes inside the cybernetic variables u_i and v_i:

$$u_i = \frac{p_i}{\sum_j p_j} \tag{6.6}$$

$$v_i = \frac{p_i}{\max_j (p_j)} \tag{6.7}$$

6.3.2 Unregulated vs. Regulated Kinetics

For the sake of analyzing the relationships between regulated and unregulated kinetics in simple terms, let us first assume that the kinetics themselves are a simple linear form with only one substrate, s, being consumed

$$r_i^{kin} = k_i s \tag{6.8}$$

Note that other factors such as saturation and enzyme concentration are presently ignored. To establish how fitting affects parameters in the model, it is necessary to further probe the regulated kinetics.

$$r_i^{reg} = v_i r_i^{kin} \tag{6.9}$$

The regulated kinetics ultimately determine changes in metabolite concentrations and need to be established accurately via optimization for the model to accurately fit the system's data. Expanding r_i^{reg}:

$$r_i^{reg} = \frac{p_i}{\max p_j} r_i^{kin} = \frac{w_i r_i^{kin}}{w_j^{\max} r_j^{kin,\max}} r_i^{kin} = \frac{w_i k_i s}{w_j^{\max} k_j^{\max} s} k_i s = \frac{w_i k_i^2}{w_j^{\max} k_j^{\max}} s \tag{6.10}$$

and truncating all of the parameters into a single parameter, θ_i, the regulated rate becomes

$$r_i^{reg} = \theta_i s. \tag{6.11}$$

Taking this perspective of the combinations of weights and kinetic parameters, the parameter set $\mathbf{k} = \{k_1, ..., k_n\}$ will be fit accordingly so that the combined parameter set θ, will provide regulated kinetics that are faithful to the data given any arbitrary selection of the weight set $\mathbf{w} = \{w_1, ..., w_n\}$. For systems with Michaelis-Menten kinetics, a similar result comes about assuming that the Michaelis-Menten constants, K are all the same for each pathway

$$\begin{aligned} r_i^{reg} = \frac{p_i}{\max p_j} r_i^{kin} = \frac{w_i r_i^{kin}}{w_j^{\max} r_j^{kin,\max}} r_i^{kin} = \frac{w_i \frac{k_i s}{K+s}}{w_j^{\max} \frac{k_j^{\max} s}{K+s}} k_i \frac{k_i s}{K+s} \\ = \frac{w_i k_i^2 s}{w_j^{\max} k_j^{\max} (K+s)} = \frac{\theta_i s}{K+s}. \end{aligned} \tag{6.12}$$

From the above, it is clear that the weighting factors in the u_i and v_i expressions are balanced out by the choice of parameters $\mathbf{k}$. This helps to explain the phenomena in the preceding section where multiple objective functions provided feasible gene expression predictions.

6.4 Cross-Validation Study of Regulatory Weights

In the last two sections, it has been made clear that the choice of weights, those representative of each pathway's ROI, for the cybernetic model do not significantly influence the ability of the model to qualitatively represent gene expression. These were approximated by the control variable for induced enzyme production, u_i.

To further probe the question "How do we validate the weights and the objective question they represent?", a 3-dimensional weight space was sampled for each of the three pathways that represents an arbitrary set of potential objective functions. These weights were then used to fit parameters for the unregulated kinetics in a model that describes the formation of prostaglandins (PGs). Given that the weights appear in both the numerators and denominators of both cybernetic expressions in the models, their values were constrained between 0 and 1. This constraint, will not change the result of this analysis as the weightings between different pathways can be fully represented in this interval. To visualize the weight set used in this analysis, the 3d scatterplot in figure 6.3 shows the spatial distribution of weights used in the different models.

The blue dots represent each weight set analyzed and its corresponding weight value, $\mathbf{w} = (w_{PGE_2}, w_{PGF_{2a}}, w_{PGD_2},)$, for each PG pathway. What should primarily be observed in the 3-dimensional diagram is that the values of the randomly sampled weight values evenly cover the space that we are trying to analyze. Parameters were fit using a time-course set of PG data for four total conditions. These were: a control; a compactin treatment, a statin drug that perturbs PG formation in addition to inhibiting cholesterol formation through HMG-CoA Reductase; a KLA treatment

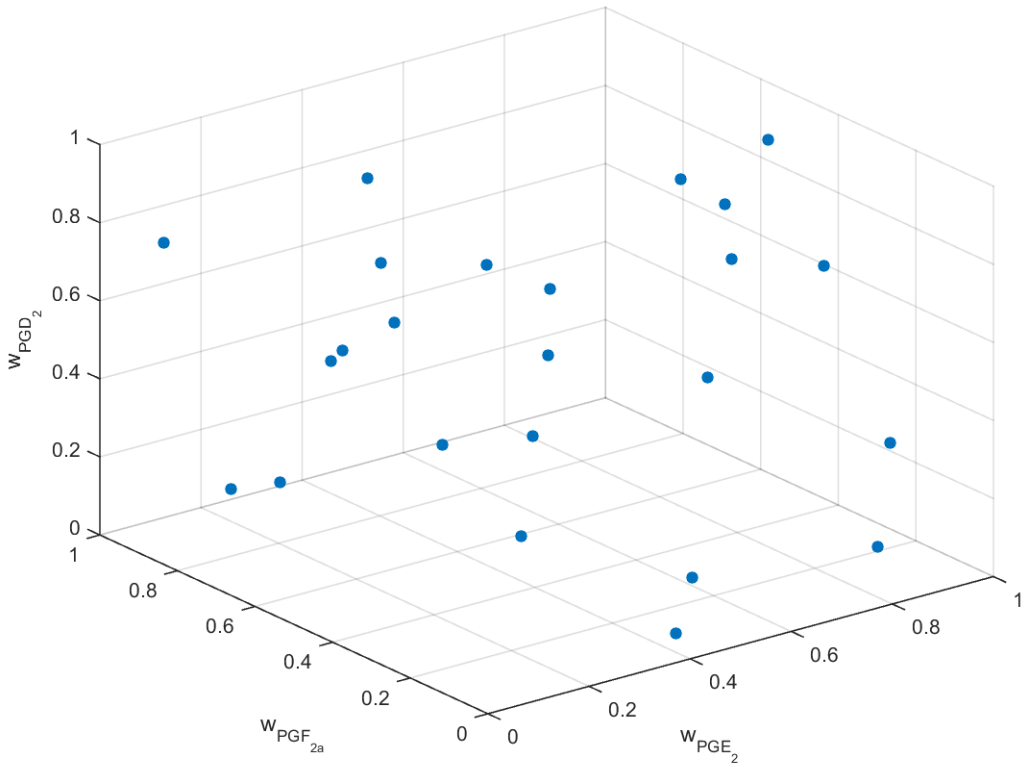

Figure 6.3.: Scatter plot of weights tested representative of arbitrary objective functions. Randomly sampled weights are plotted with each axis representing weight value for one of three pathways.

which stimulates PG formation as an indicator of present infection; and a combined compactin and KLA treatment.

To allow for cross-validation of the model, only 3 of 4 data sets were used for the four conditions under which the data were taken. The combined compactin and KLA treatment were left out to analyze the robustness of each model's set of unregulated kinetics. To fit all 16 parameters in the model to the data for the 24 trials, a numerical optimization method was applied that employs a genetic algorithm, starting with a randomly seeded initial population, followed by a gradient search. The function that was minimized was the sum of squared errors of the model for the different conditions.

25 random starting points were surveyed in the parameter space for all 24 sets of weights. This survey of the parameter space was by no means computationally trivial. With an average time of 15 minutes per parameter set, total sampling of all 600 parameter sets took roughly one week on a desktop computer (Intel Core i7-2600). A histogram showing the minimization values of the SSE for the best parameter set for each $\mathbf{w}_i$ is in figure 6.4.

To provide a more complete sense of what these SSE values mean, after running this minimization scenario for the original weights for TNF-alpha for a total of 2000 initial parameter values, the minimum SSE is just under a value of 5. A visualization of the fit of a model with a mean value of these SSEs above, observe the fit of a model with an SSE value of 8.14 in figure 6.5.

As is evident in figure 6.5, most all weight sets provide an adequate fit of the metabolite data modeled. This provides some insurance that the fitted parameters in the foregoing analysis are valid.

Moving on to the cross-validation of the different potential weightings using the KLA+compactin data, let us first visualize the fit of this treatment for the whole set of models in figure 6.6.

As is seen in figure 6.6, some of the weighted fits leave much to be desired for the cross validation set. Others, provide results that are similar to the KLA treatment.

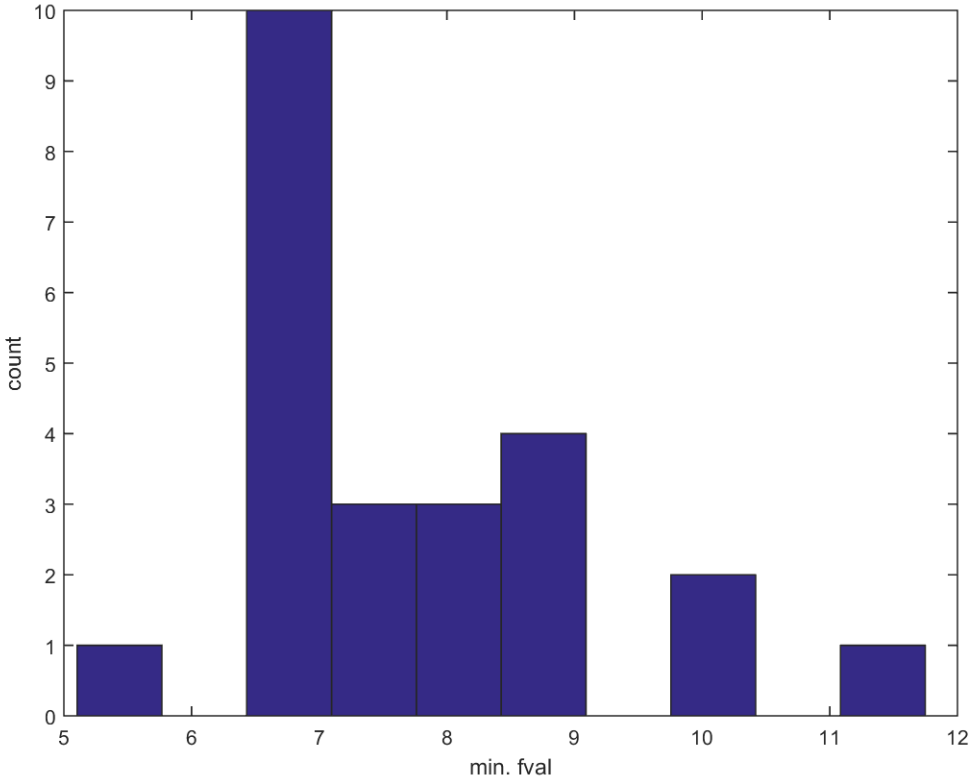

Figure 6.4.: Histogram of SSE scores for parameters determined using arbitrary objective functions.

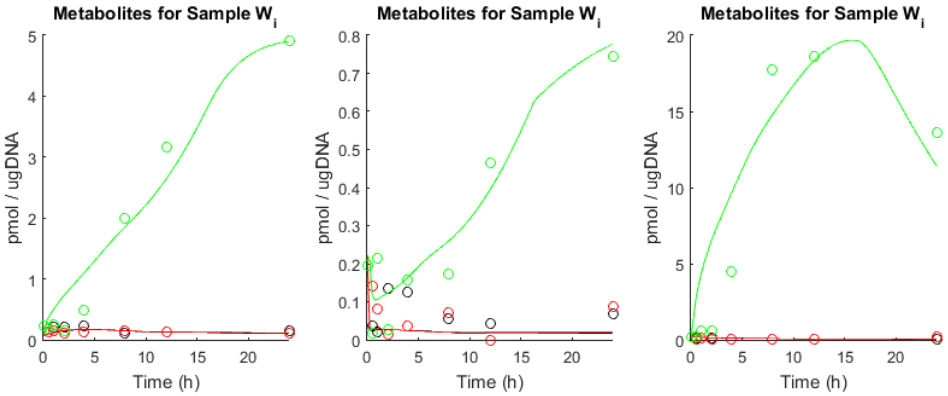

Figure 6.5.: Visualization of model fit with SSE score of 8.14.

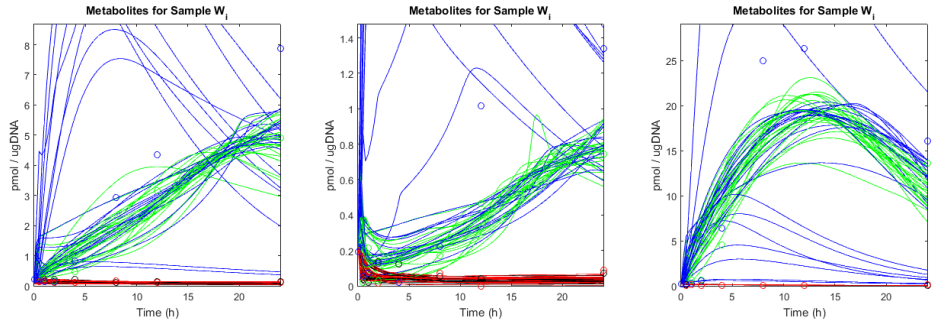

Figure 6.6.: Visualization of fit of all models. The control scenario is in black, compactin treatment in red, KLA treatment in green and the CV set (KLA+compactin) is in blue.

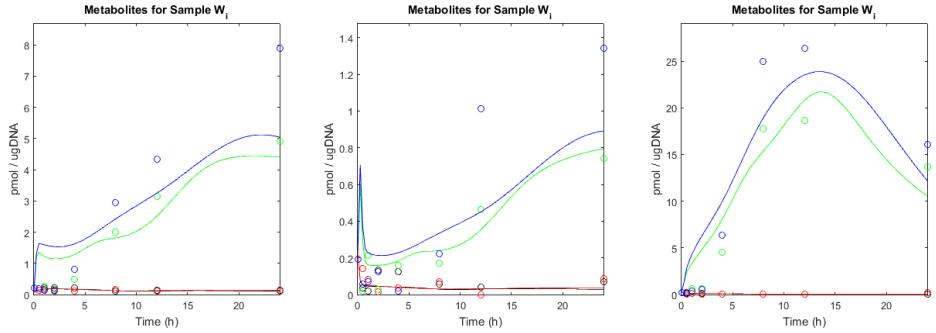

Figure 6.7.: Visualization of fit of original TNF-alpha objective function model.

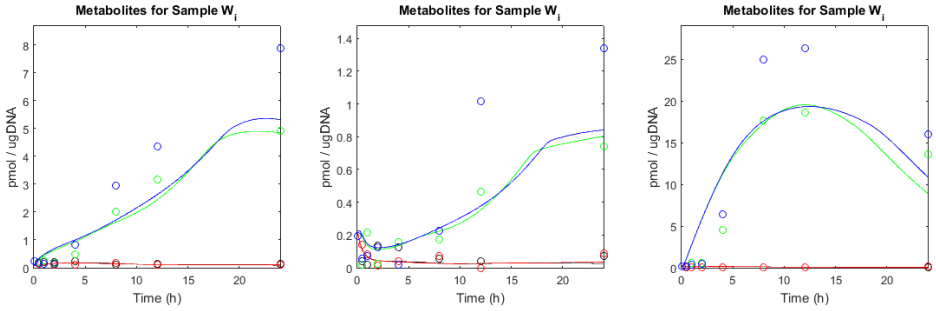

Figure 6.8.: Fit of the best cross-validation weightings from the randomly sampled objective functions.

The original TNF-alpha objective function provides the fit of the cross validation data set in figure 6.7.

Comparing the fit of the TNF-alpha objective function in figure 6.7 with the best cross-validation fit from the random sampling of weights in figure 6.8, it appears that the TNF-alpha weights provides better qualitative results in that the KLA + Compactin case is above the KLA.

6.4.1 A Note on KLA + Compactin Treatment

Before jumping to the conclusion that neither model is descriptive of the KLA + compactin scenario, an import point should be clarified. The results of KLA + com-

pactin treatment (henceforth referred to as the combined treatment) are somewhat counterintuitive. Relative to the KLA treatment, the combined treatment reduces the level of this system's input lipid, arachidonic acid or AA. Despite this decrease in the concentration of the input, the concentrations of products of the metabolic network increase. This effect of compactin treatment, in addition to KLA, is hard to gauge when looking at the compactin data alone; the compactin and control scenarios yield similar magnitudes of output prostaglandins. Therefore, it would be difficult for any model to accurately capture the exact result of the combined treatment which was, itself, a notable find experimentally [89]. In spite of the fact that the TNF-alpha weights do not perfectly model the cross-validation set, the fact that they are able to predict this sophisticated increase in PG level to any degree is an endorsement of the model's unregulated kinetics and objective function.

6.4.2 Deeper Analysis of Objective Weights

Even though the TNF-alpha model weights do seem to provide a more accurate description of the cross-validation set, it is hard to say with certainty that this is absolutely true. Unfortunately, more experimental data for this system is unavailable to verify the unregulated kinetics of the various weights studied. Even though the kinetics themselves cannot be further validated in this absence of experimental data, further analysis of the weights can still be performed. Many relevant questions to the weights still remain such as: If we look at the spatial distribution of weights and their ability to fit the cross-validation data set, do any discernible trends arise? And, furthermore, do the values of weights correlate with the ability of a model to fit data?

The spatial analysis of weights mentioned above was performed in the following way. The weights were plotted 3-dimensionally as in figure 6.3 above. Below in figure 6.9, the points representing the locations of the weights are sized according each weight set's SSE for the cross-validation set. This means that larger points have larger SSEs than smaller points. Also, the color is representative of the SSE value

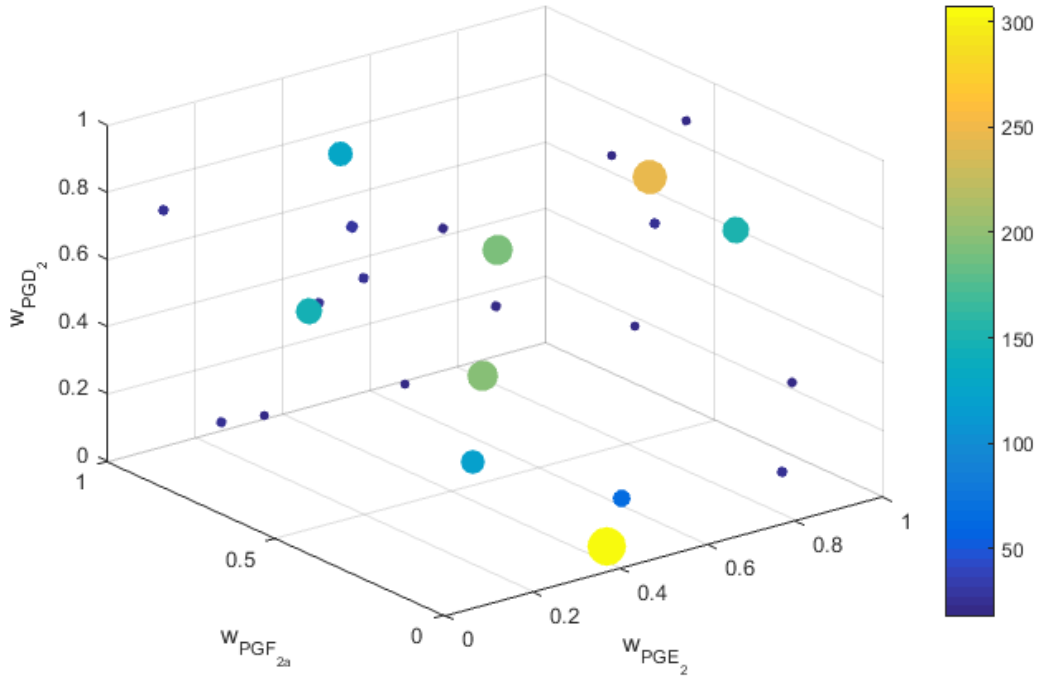

Figure 6.9.: Spatial representation of cross-validation SSEs for randomly sampled objective function weights in 3-dimensions.

with purple representing more accurate cross-validation predictions than blue, green and yellow.

In figure 6.9, it appears that there is no obvious relationship between different regions in the weight space and the values of the cross-validation SSE. The small purple dots are evenly distributed throughout the space and the larger dots are mixed evenly with the larger dots. To further analyze this, observe the depiction in figure 6.10 of the correlational matrix for a set of values containing the weights for PGE_2, PGF_{2a}, and PGD_2 in the first three columns and the SSE values for the CV set in the forth. In figure 6.10, the panels of interest are in a red box.

Pearson's correlation coefficients for the various scenarios are noted in the upper left corner of each scatter plot. It is clear from this analysis that the values of the

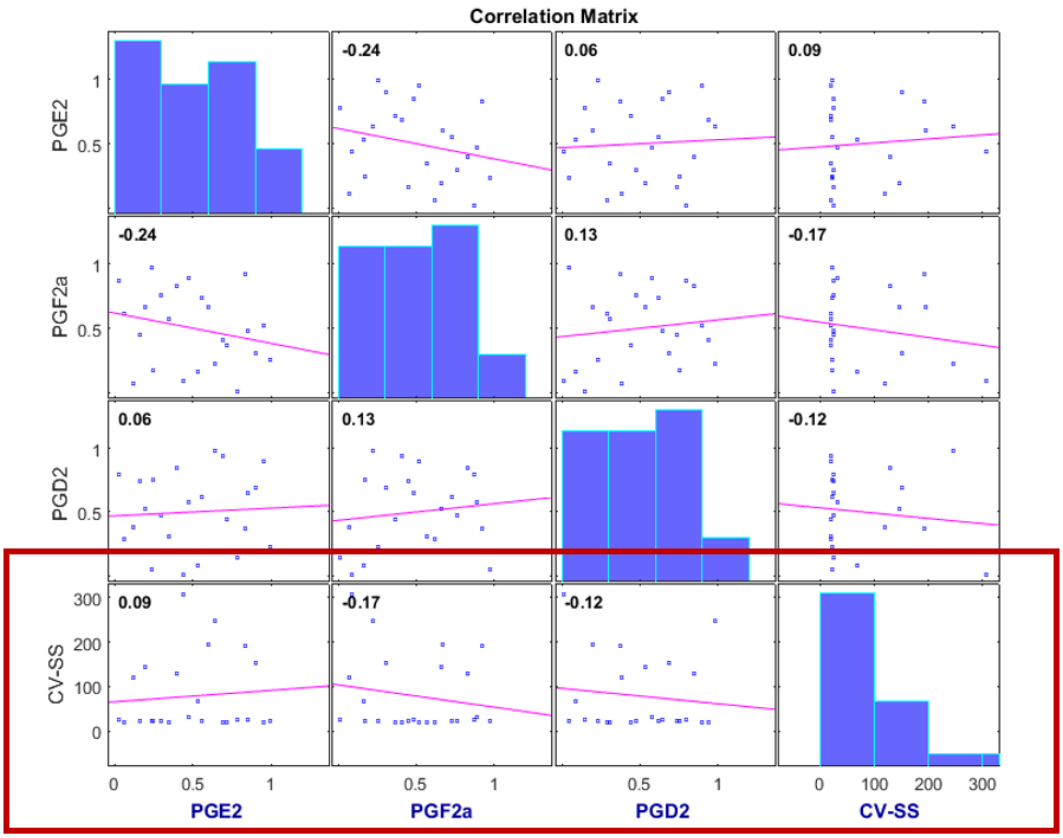

Figure 6.10.: Visualization of correlation matrix for various weightings with cross-validation SSE values. Plots in red box show correlation between various weight values and cross-validation SSE.

weights do not have any significant relationship with the ability of the model to predict the objective function. Both of these methods of scrutinizing the weights of the various models paint a similar picture. While the TNF-alpha function allows for the best fit of the cross-validation data, there is no discernible relationship between the values of the weights and their ability to predict the cross-validation case.

This implies either one of two things, that there is some singular point in the weight space that provides the best result. In this case, the singular point could possibly be the TNF-alpha weights. The other, is that there is no apparent relationship between the choice of objective function - in the form of weights - and the ability of the model to fit the data.

6.5 Discussion

6.5.1 Drawbacks Associated with Parameter Estimation

This analysis suffers from one significant drawback which is the nature of nonlinear parameter estimation. Because the parameters for this system are unknown, it is hard to conclusively say that the results of nonlinear parameter estimation have reached a global minima in the error function. Even for this relatively simple system with five metabolites and 3 modeled enzyme quantities there are a total of 16 parameters. Trying to find a minimum in this parameter space is difficult and can only be found numerically. Ideally, the search for parameters would involve some kind of smooth surface with an absence of local minima. To illustrate this, consider the surface of SSE values (z-axis) in Figure 6.11 for the model for two parameters (the x and y axes) in the model and no variation of the other parameters.

The minimum of this low-dimensional surface can be found using numerical methods. However, the complexity of this search is much higher in that while searching for a minimum value of the SSE function for these two parameters, the minimization algorithm also has to optimize the values an additional 14 parameters. To better illustrate the minimization of SSE for these two parameter values, consider the surface

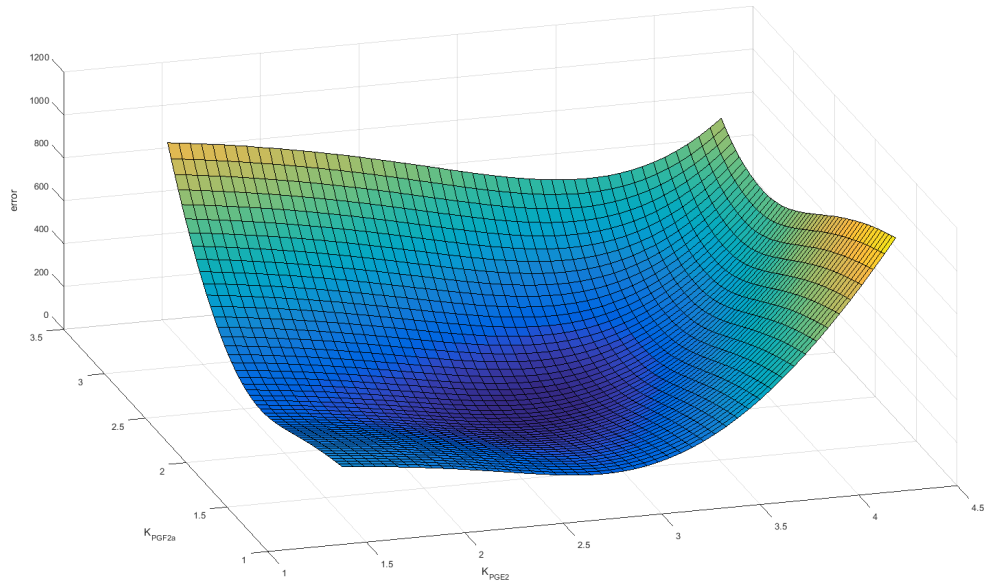

Figure 6.11.: Visualization of surface of error function used for parameter estimation for two kinetic parameters in the model without variation in other parameter values.

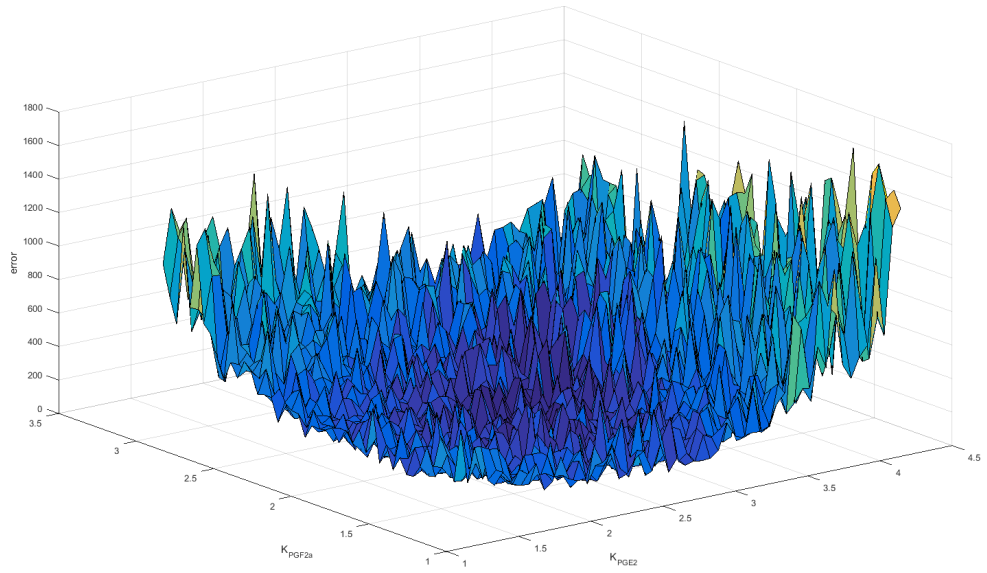

Figure 6.12.: Visualization of surface of error function used for parameter estimation for two kinetic parameters in the model with 5% variation in other parameter values.

(figure 6.12) in the case where the values for all other parameters change arbitrarily within an interval 5% of some informed guess at the correct value for the parameter. The surface representative of this search space is extremely rough pockmarked by a multitude of local minima.

Provided that the true parameters for each set of weights cannot be found, one faces a dilemma. Even if we sample the parameter space more completely, we will never be certain that the parameter values are a truthful representation of the underlying unregulated kinetics. At the same time, with this incompleteness in the parameters, it is difficult to say that these regulatory weights are important in providing the model with an ability to fit the kinetics. The analysis of matching law control variables indicates that the weights are not important when it comes to predicting the general behavior of mRNA creation. However, the weights towards TNF-alpha do show a slightly better fit on the cross-validation data. This could be an argument

for the initially selected objective function. The main snag in this reasoning is the apparent lack of correlation between weight values and the ability of a model to fit the combined treatment scenario. As mentioned before, there could be a singular point in the weight space that allows for the best cross-validation of the model, but one could never determine that without a high confidence in the parameters used in the analysis.

6.5.2 Final Thoughts

While many weights are able to fit the gene expression data, there was no discernible trend among weights to predict the cross-validation case. This issue is central to the formulation of cybernetic models in that one has to develop a cybernetic model upon a central postulate which is the objective function. In the case of cybernetic modeling of bacterial cells, dynamic objective functions like growth rate maximization or substrate uptake rate maximization make sense from an evolutionary point of view. When shifting focus to the modeling of dynamic regulation in complex mammalian systems, postulation of an objective function is not straightforward. In the prior case-study of PG metabolism that spans this chapter and the last, it is evident that many objective functions will work. Because of this, there is a demonstrable need for a refined approach at establishing objective functions.

On one hand, it is feasible that one merely makes a clear statement that the objective function is a postulate central to the model developed. Specifically, there is no need to go further than explicitly stating that the objective function is a central assumption of the model. On the other hand, it would be better to analyze biological data in a structured fashion to make conclusions about what the objective function is based off of statistical analysis and information gathered from curated biological databases. A novel method to do this is proposed in the next chapter.

To make this analysis more complete, a better method to estimate parameters that avoids complex nonlinear optimization is then proposed in the chapter 8.

7. LEARNING CYBERNETIC OBJECTIVES THROUGH OBJECTIVE ANALYSIS OF OMIC DATA

Summary

Cybernetic models describe metabolic changes by means of simplifying intricate cellular regulation mechanisms with the notion of dynamic metabolic goals. In prior work, these goals have taken the form of inductive assumptions about the nature of metabolism and are explicitly stated as objectives like "maximize the rate of growth" or "maximize the rate of carbon uptake." These inductive metabolic goals have yielded numerous predictions of metabolic phenomena, but a procedure to determine the most descriptive metabolic goal for a given system has yet to be proposed. To address this, a method to deductively resolve metabolic objectives from a combination of metabolomics and transcriptomics is developed. This analysis considers both metabolic changes and the regulation of enzymes that influences those metabolic changes. It provides an objective function that takes the form of optimal weightings that represent the returns on investment for metabolic pathways that are competing for resources. To give biological context to these returns on investment, a method to mine gene expression data and explain the objective function using pathway enrichment analysis is presented. This approach is used to analyze three systems: two diauxic growth scenarios and prostaglandin metabolism in a mammalian cell line. Pathway enrichment analysis of the genes mined using this procedure provides objective functions that show agreement with already established experimental knowledge related to the behavior of these systems in their respective conditions. The ability to determine objective functions deductively from the data provides for robust cybernetic descriptions of systems where the objective function is difficult to determine.

7.1 Introduction

For any metabolic process, there are a number of objective functions that are adequately descriptive. The common approach for modeling bacterial systems through the use of objective functions is to assume that the cultured cells are regulating their metabolic processes towards the maximization of growth yield [33] or rate [3]. A popular objective function that has provided for the cybernetic prediction of unique metabolic phenomena is the maximization of carbon uptake rate [20].

Overall, the utility of objective functions in cybernetic models is the compression of many individual regulatory mechanisms of biological control into an overarching principle of regulation. In single-celled organisms, objective functions that properly represent regulation of metabolism towards states that maximize a biological entity's ability to propagate its genes are adequately representative of teleonomic principles [106]. Cultures of bacteria, typically modeled on the scale of numerous generations, must compete fiercely to ensure the survival of their genes and cybernetic objectives like the maximization of growth rate or resource consumption have intuitive validity.

Prior work has relied on the inductive assumption of an objective which is validated through the examination of a model's ability to predict different metabolic phenomena [9, 21, 25, 26]. While successful, this approach ignores a multitude of potentially descriptive statements of a particular system's metabolic goal. This is especially relevant in systems where the objective function is not as clearly defined as it is in single-cellular organisms. For example, cells in multicellular organisms cooperate and may have different objective functions that depend on the nature of the cell of interest. It is possible that hepatocytes have different objective functions than adipocytes. The objectives of these two cells may even be complimentary in some way. For these systems, inductive assumption of an objective function is possible. However, the ability to deductively assert an objective function for a system has a greater logical strength than a modeler's prima facie assertions of what an objective function *ought to be* given what is understood about the cells. In light of this,

the present work attempts to address the objective function identification problem by employing methods to analyze dynamic cellular data. Other attempts have been made at deductively asserting an objective function's value for constraint-based models [48, 107, 108]. These attempts are characterized by a common limitation; they do not make use of regulatory information and focus on models that employ yield-based as opposed to rate-based optimization. The present work, on the other hand, seeks to identify objective functions through the mining of data that is representative of dynamic cellular regulation.

The similarity between the cybernetic variable u_i, representative of the control of induced enzyme synthesis, and gene expression has been demonstrated in chapters 4 and 5. Using this notion, an objective function can be deduced through the simultaneous analysis of time series metabolite and gene expression data. The similarity between these two sets can be used to ascertain a deeper understanding of regulation and generate what is termed a "goal signal." The goal signal is then analyzed further in the context of gene expression profiles, pathway knowledge and already established biostatistical methodologies. At a high level, the method proposed mines the entirety of time-series gene expression data to determine a set of similar genes that are representative of the objective function's goal signal. The ability to extract objective functions from biological data provides for a robust description of metabolic regulation. Furthermore, it endows the modeler with greater confidence in the robustness of these predictions for systems as it can explicitly identify metabolic goals where the objective function is not easily induced. Lastly, insight into the metabolic goals of metabolic systems allows for a more holistic understanding of how the states of cells evolve over time. The applications of this approach are as many as there are systems that can be modeled with cybernetics.

7.2 Methodology

This approach consists of five steps. The first calculates optimal return on investments for competing metabolic pathways by calculating regulation that optimally fits

the metabolic network's genes. This is followed by generating a summary goal signal for the metabolic network. The goal signal is compared with the total picture of cellular regulation to generate a list of genes that show statistically significant similarity to the goal signal. The gene list is then analyzed using pathway enrichment analysis to identify pathways that are overrepresented in the gene list. The literature is then consulted to identify links between the overrepresented pathways and the metabolic system of interest. This leads to a conclusion about the system's metabolic goal. A bird's eye view of the method employed to learn objective functions from the data is depicted in figure 7.1.

7.2.1 Background: Cybernetic Modeling

Before the data driven search for metabolic objective functions is fully fleshed out, it is necessary to cover some preliminaries in the meaning of cybernetic goals. To provide the reader with a sense of how cybernetic models are formulated, a simpler version of a cybernetic model will be explained.

In cybernetic modeling, there is a dichotomy between so-called "regulated rates," r_i, and "unregulated rates," r_i^{kin}. The unregulated rates are the enzyme-specific kinetics that determine the maximal rate at which a reaction or pathway can be expressed. The formulation of the unregulated kinetics is not fixed to one particular function of substrate concentration, but as a point of reference, they frequently take a Michaelis-Menten form such as

$$r_i^{kin} = \frac{e_i k_i s}{K_i + s}, \tag{7.1}$$

where e_i is the enzyme level, s is the concentration of substrate, and k_i and K_i are rate parameters for the i^{th} pathway. The regulated rates are a function of the unregulated rates and determine the time dependent changes in metabolites, $\frac{d\mathbf{x}}{dt}$. From the cybernetic viewpoint, activation of pathways is proportional to the return on investment (ROI) that a pathway provides [11]. For instance, in diauxic growth, a substrate's rate of growth could be the ROI and the pathway for the substrate with

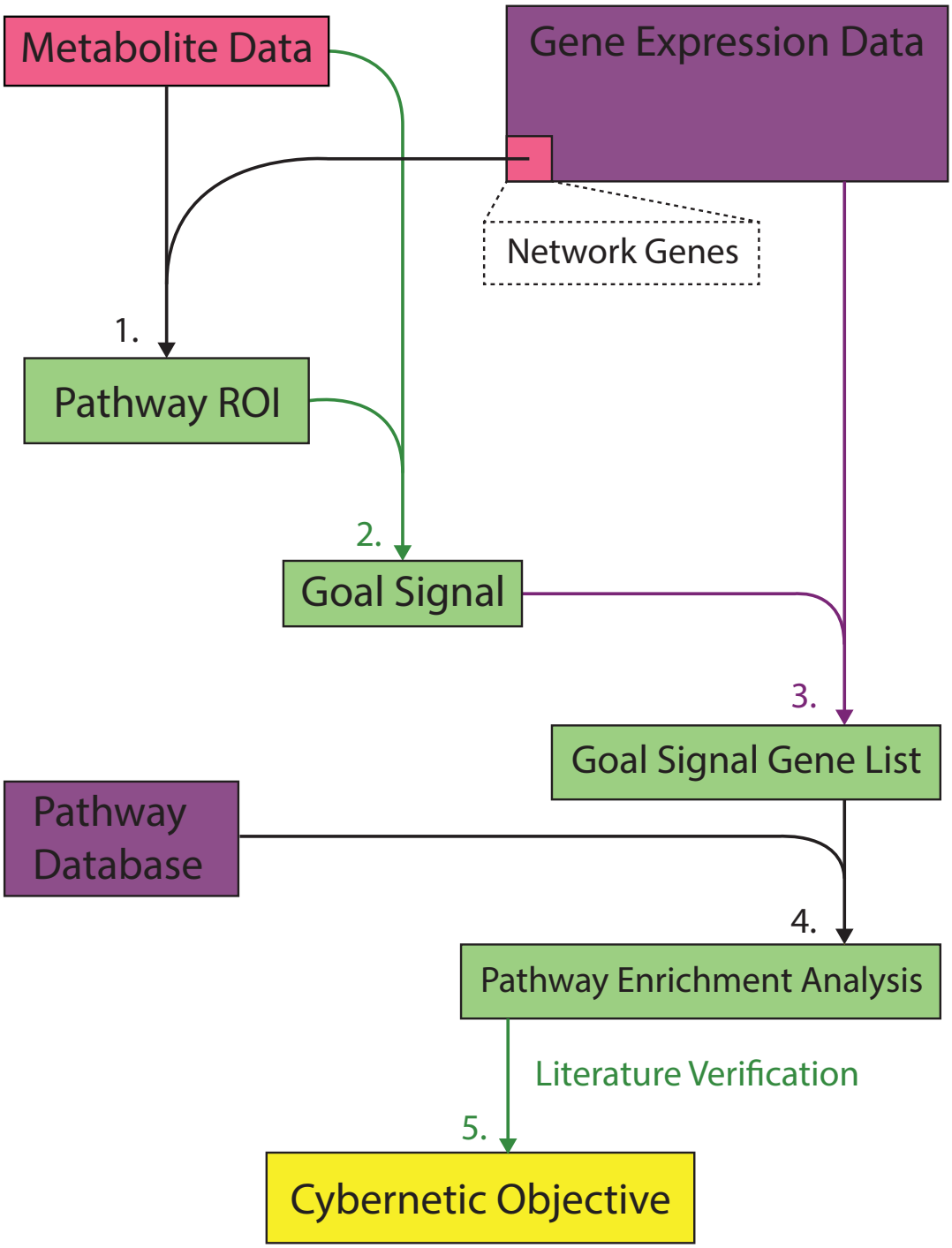

Figure 7.1.: Overview of the use of data to analyze objective functions.

the maximum ROI is fully activated while submaximal ROI pathways are proportionally turned down. The regulated kinetics are determined by the product of the proportional activation cybernetic variable v_i with the unregulated rate as

$$r_i = v_i r_i^{kin}, \tag{7.2}$$

with v_i being a function of ROI which is represented by p_i, for each competing reaction or pathway in the metabolic network. This proportional law takes the form

$$v_i = \frac{p_i}{\max\limits_j (p_j)} \tag{7.3}$$

The other key feature of cybernetic models is the approximation of changes in enzyme levels, e_i:

$$\frac{de_i}{dt} = \alpha + u_i \frac{k_E s_i}{K_{E,i} + s_i} - (\mu + \beta) e_i. \tag{7.4}$$

This expression for change in enzyme level includes α for the constitutive generation of enzymes, as well as μ and β for the depletion of enzymes due to growth dilution and degradation respectively. The remaining term is for the induced expression of enzyme and is a function of substrate concentration and cybernetic variable u_i. This control variable regulates induced enzyme formation according to the ROI relationship:

$$u_i = \frac{p_i}{\sum_j p_j} \tag{7.5}$$

This control law states that the fractional investment of the finite pool of cellular resources for enzyme production among competing pathways will be proportional to the ROI gained by investment. Prior work in chapters 4 and 5 has demonstrated a convincing correlation between the cybernetic variable for induced enzyme synthesis control and data representative of mRNA levels. While the u_i variables themselves cannot be directly substituted for gene expression, they are representative of control over the same process, enzyme formation. Given this functional similarity, their scaled versions should change in approximately the same fashion.

The ROI for the metabolic network is subject to the objective function's formulation. The formulation of ROI is assumed to take the form

$$p_i = w_i r_i^{kin}, \tag{7.6}$$

where w_i is some weight that states which pathways have a stronger relationship with the metabolic goal than others. As an example, these weights in prior work have been carbon number for the maximization of carbon uptake rate [20]. In this work, the weight will be assumed to be derived from the following relationship to the metabolic goal, J, that is expressed by

$$\dot{J} = \sum\nolimits_i w_i \dot{x}_{p,i}, \tag{7.7}$$

where $\dot{x}_{p,i}$ represents the time-dependent change in some product of the metabolic network. The changing metabolic goal for the system is a function of weighted rates of product formation from the metabolic system and the combination of weighted product formation rates makes up goal signal profile, $\dot{J}(t)$. The purpose of the goal signal is to compress the information generated by the metabolic network into a single profile that changes over time. This work first seeks to establish what the best weights are through the treatment of metabolite data and gene expression data. Subsequently, statistical analysis of the whole set of gene transcript levels and their relationship to weighted metabolite data works to resolve the meaning of the objective function J in biological terms.

7.2.2 Data Requirements

The method to establish objective functions deductively first starts with dynamic biological data. The requisite information to perform this analysis consists of two dynamic sets of data taken for the same system under the same set of conditions: metabolite levels and gene expression data. Expansive gene expression data sets allow for a comprehensive analysis of possible objective functions. Because of this, microarray series are used in the following analyses. Other data sets that are representative of gene expression such as RNA-seq can also be used but were not available for the systems analyzed.

In this work, three test cases are examined. Two focus on diauxic growth of *E. coli* in two separate substrate mixtures. The first of these consists of growth on glucose followed by growth on acetate and is developed using data from [65]. The second

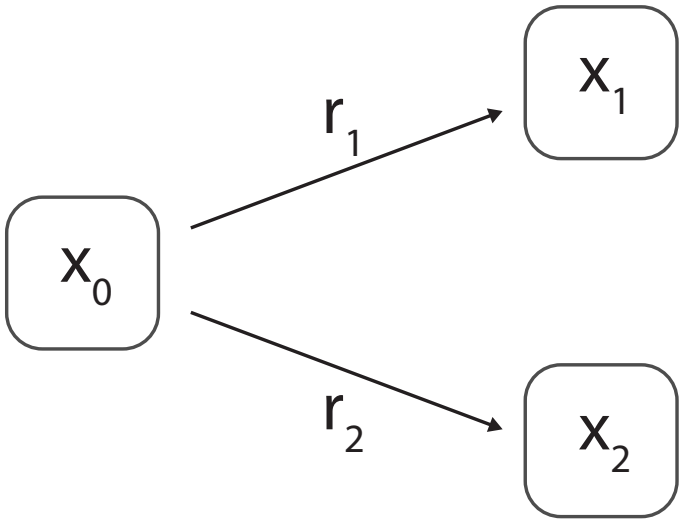

Figure 7.2.: Toy example of metabolic network. x_i represent metabolites and r_i represent regulated reaction fluxes for competing pathways.

diauxic growth scenario is comprised of growth on glucose and lactose and uses data from [66]. The final test case seeks to expand this method to a more complex system prostaglandin formation in a macrophage cell line [87].

7.2.3 Determination of Regulated Rates from Data

To make the analysis more concrete, consider a toy metabolic network in figure 7.2 with three metabolites (x_0, x_1, x_2) and two regulated reaction fluxes (r_1, r_2).

Cybernetic analysis of a given dataset starts with approximation of the metabolic fluxes through different pathways. The metabolic fluxes for a set of reactions or pathways can be estimated from the following statement where the change in metabolites normalized by cell concentration is roughly equal to the flux through the reaction that generates it:

$$\frac{1}{c}\frac{dx_i}{dt} = r_i. \tag{7.8}$$

The changes in metabolites above can be estimated through finite difference methods by estimating the slope between the data points. If there is some downstream process that consumes either x_1 or x_2, the flux can be modeled using a linear decay process such as

$$\frac{dx_i}{dt} = r_i - \gamma_i x_i \tag{7.9}$$

where the decay constant, γ_i, is taken from some literature value, or approximated by means of model fitting. In the case of an irreversible reaction, this decay constant can be estimated crudely via forcing the minimum reaction rate to be some feasible value greater than zero such that

$$\min_t(r_{i,t}) > 0 \tag{7.10}$$

for some time t over which the data is taken. More exhaustive approaches to estimating fluxes from metabolite data are available [109, 110]. The estimation for pathway fluxes for more advanced pathway-based cybernetic models is detailed in the section for this method's generalization to hybrid cybernetic models.

7.2.4 Cybernetic Objectives

The aim of this work is to deductively ascertain the objective function for a given metabolic network. For some metabolic network generating metabolic products, the products can be viewed as a means that the organism uses for achieving an objective. The weights in the above representations of ROI are a channel through which metabolic products can be connected with the goal J. In situations where the goal is not known or easily inferred, the calculation of the most appropriate weights for the system is possible. This method relies on the computation of u_i profiles from metabolite data and then fitting weights to optimize the fit between experimental gene expression data that is representative of the control over reaction or pathway catalyst formation. These weights determined from this fitting procedure are termed the optimal weights. The calculation of u_i variables from an arbitrary set of objective weights achieves this. To calculate these control variables from the data, the unregulated rates

are determined from the observed regulated fluxes. These are estimated through the manipulation of the relationship between regulated and unregulated fluxes and the foregoing definition of ROI ($p_i = w_i r_i^{kin}$):

$$r_i = v_i r_i^{kin} \tag{7.11}$$

$$r_i = \frac{p_i}{\max_j(p_j)} r_i^{kin} \tag{7.12}$$

$$r_i = \frac{w_i r_i^{kin}}{\max_j(w_j r_j^{kin})} r_i^{kin} \tag{7.13}$$

$$r_i^{kin} = \sqrt{\frac{r_i \max_j(w_j r_j)}{w_i}} \tag{7.14}$$

The maximum weighted rate of the set of competing metabolic weights at time t is simply the maximum regulated rate given the property that $v_i = 1$ for this pathway. Mathematically:

$$\max_i(w_i r_i) = \max_j(w_j r_j^{kin}) \tag{7.15}$$

With r_i^{kin} known for each pathway, the u_i control variables for each pathway are fixed for any arbitrary set of weights w_i

$$u_i = \frac{w_i r_i^{kin}}{\sum_j w_j r_j^{kin}} \tag{7.16}$$

Provided that the u_i variables can be calculated for any set of weights given the metabolic data which allows for the estimation of the regulated rates, the optimal set of weights $\mathbf{w}_{opt}$ can be calculated where

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \left(\sum_{t=1}^{n} \sum_{i=1}^{m} \left(u'_{i,t}(w_i) - g'_{i,t}\right)^2\right), \tag{7.17}$$

in which there are m competing pathways and n data points. Gene expression values for the catalyst relevant to pathway or reaction i are denoted by g_i. The prime on the gene expression and control variable represent the scaling of both of these quantities for their qualitative comparison. This scaling for each pathway or reaction over a series of time is

$$u'_i = \frac{u_i - \mu(u_i)}{\sigma(u_i)} \tag{7.18}$$

The scaling of gene expression takes the same form. Both are necessary to make for proper numerical comparison of both quantities as the magnitudes of g_i vary considerably. The scaling of both variables highlights the fact that it is difficult to make direct comparison of the absolute magnitudes of the two quantities. For pathways where there are multiple genes that describe thei^{th} pathway's behavior, one can generalize the pathway gene as a mean of the sets total changes in gene expression as

$$g_i \approx \mu(g_k). \tag{7.19}$$

This formulation ignores the strength of contribution of each reaction to the pathway but is convenient to use in situations where the complete metabolic pathway is incomplete (e.g. in lumped cybernetic models). The generalization of this approach to hybrid cybernetic models [19] is in the following section. Included is a more rigorous way to approach to approximate pathway specific gene expression.

The vector of weights $\mathbf{w}_{opt}$ for the set of reactions or pathways encapsulates the optimal objective function to best fit both metabolite and gene expression data. Given the ability of these weights to best estimate both sets of data, they are assumed to capture the best objective function for the metabolic data and will provide for robust prediction of different metabolic states.

7.2.5 Generalization to Hybrid Cybernetic Models

The determination of fluxes can be generalized to larger systems relevant to hybrid cybernetic models. These models use elementary modes to approximate reaction fluxes and compare the rates of uptake through these competing metabolic pathways. The uptake fluxes, $\mathbf{r_M}$, have the following relationship with the reaction rate fluxes, $\mathbf{r}$, as:

$$\mathbf{r} = \mathbf{Z}\mathbf{r_M}, \tag{7.20}$$

where the network $\mathbf{S}$ has the elementary flux modes (EFMs) $\mathbf{Z}$, a comprehensive representation of the pathways/macroscopic reactions through the metabolic network

[18]. The change in measured metabolite vector $\mathbf{x}$ of n metabolites is modeled by the following:

$$\frac{1}{c}\frac{d\mathbf{x}}{dt} = \mathbf{S_x Z V r_M^{kin}}. \tag{7.21}$$

where $\mathbf{S_x}$ is the stoichiometric matrix associated with only metabolites that are measurable and $\mathbf{V}$ is a diagonal matrix containing proportional control variables for each pathway v_i. For more detail on hybrid cybernetic models, refer to [19]. If one selects n elementary flux modes from $\mathbf{Z}$ that best represent the span of the flux hypercone defined by $\mathbf{Z}$ [22], then the unregulated uptake fluxes, $\mathbf{r_M}$, are determined in a similar fashion to the above analysis stemming from the useful property that the product of $\mathbf{S_x Z}$ is invertible. Removing the term for uptake flux regulation, we have

$$\mathbf{r_M} = \frac{1}{c}[\mathbf{S_x Z}]^{-1}\frac{d\mathbf{x}}{dt}. \tag{7.22}$$

These are the regulated uptake fluxes for each elementary mode. The relationship between the regulated and unregulated uptake fluxes is

$$\mathbf{r_M} = \mathbf{V r_M^{kin}} \tag{7.23}$$

In a similar fashion, the unregulated rates are approximated by assuming that the maximum weighted $w_i r_{M,i}$ is the same as the maximum unregulated rate $w_i r_i^{kin}$. With the maximum rate known, the known, each individual rate is then determined as it is in equations 7.14 and 7.15. The control variables for induced enzymes synthesis are determined from here by the same procedure.

To calculate the optimal set of weights, genes representative of each EFM must be selected. This is potentially difficult as EFMs can overlap in a number of the reaction pathways that they go through. A composite gene profile that is representative of all the genes in a given profile can be estimated as the sum of the products of an elementary mode's value for each reaction times the gene corresponding to that reaction's enzyme or:

$$\zeta_i = \sum_{j=1}^{m} Z_{i,j} g_j. \tag{7.24}$$

With the composite gene expression profile for each EFM, the optimal weights for the cybernetic objective function are calculated by minimizing the sum of squared errors between the u_i variables which are directly inferred from the data and the composite gene expression profiles, ζ_i.

The optimal weights can then be used to estimate a goal signal for the metabolic system which can then be used in the same way as above to guess at the biological objective function. This method can also be used to extrapolate the fluxes for lumped hybrid cybernetic models [24].

7.2.6 Further Analysis of Optimal Weights

While the weights may fit the data, they are, at this point, still a mathematical abstraction of the goal of the process. They do not have the same intuitive meaning as objective functions like maximization of growth rate. The proceeding analysis attempts to distill a biological meaning for the optimal weights using a comprehensive analysis of gene expression data. While the objective is yet to be known, the optimal weights provide some "goal signal" or time dependent profile for J. This is given by

$$\dot{J}(t) = \sum\nolimits_i w_{i,opt} \dot{x}_{p,i}(t) \tag{7.25}$$

The goal signal is a means of connecting the metabolic network's products with the optimal returns on investment. If a metabolic network is generating a set of products, from the cybernetic perspective, there is an underlying biological purpose for their generation. Metabolic products that are generated from pathways with a strong ROI contribute more to the goal signal than ones that do not. The goal signal summarizes the dynamic behavior of the metabolic network in the context of each pathway's regulation resulting from the approximated ROIs.

For some arbitrary length of time, the goal signal might go up or down depending on the network's generation of metabolites which is influenced by substrate availability and the underlying metabolic objective. This work makes the assumption that genes relevant to the organism's metabolic objective will be regulated similarly to the

network's goal signal. In other words, genes will be regulated towards goals in the same way that combinations of metabolites are produced to accomplish goals. Genes that match the goal signal will have some function related to ultimate objective of the metabolic system. This is intuitive if the genes are expressed as the product of some signaling mechanism initiated by the generation of metabolite products with negligible delay as

$$\dot{g}_{goal}(t) = \sum\nolimits_i w_{i,opt}\dot{x}_{p,i}(t) \tag{7.26}$$

The comparison of this goal signal with the entirety of gene expression data provides a list of genes that have a strong relationship with the metabolic network's goal.

To test the relationship between the goal signal and the overall picture of regulation represented in microarray data, $\dot{J}$, is compared with the numerically approximated derivatives of each gene expression profile, $\dot{g}_k$, in the set of genes. Pearson's product-moment correlation coefficient was calculated for each gene and goal signal pair $\rho_{\dot{J},\dot{g}_k}$. From this, a set of genes, G_J, associated with the goal was isolated using genes that significantly correlated with $\dot{J}$ at a p-value lower than 0.05. This threshold can be adjusted to vary the length of the gene list which should be in the range of 100-2000 genes [111].

To give biological meaning to set G_J, pathway enrichment analysis is subsequently used. Pathway enrichment analysis (PEA) compares G_J with sets of genes that make up pathways [112]. It identifies which genes in G_J are overrepresented in any biological pathway listed on KEGG [97,98]. For example, if a sufficiently high number of genes in G_J are from a certain KEGG pathway, that pathway is considered to be enriched. Given the overrepresentation of a pathway in G_J, it is considered to have some relationship with the goal. For more detailed information on how to carry out PEA, refer to the following references [111, 113].

The goal related pathways determined by PEA are then surveyed in the literature to gauge their relevance to the metabolic system. For instance, if a pathway is presented as the goal for a system under a certain condition by this analysis, a literature search using keywords for the pathway, metabolic system and condition can be made.

If there is strong evidence of a relationship of a link between the pathways selected by PEA and the metabolic system of focus, the objective determined by this method is validated.

While a pathway itself may have a relationship to the metabolic objective, an important distinction must be made. The pathways isolated by pathway enrichment analysis are feasible explanations for the biological significance of a given set of weights. These single pathways, by themselves, do not necessarily constitute a single objective function, but rather, the combination of enriched pathways provide a picture of processes related to the metabolic objective function. While not as clear as objective functions like "maximize rate of growth," these pathways provide a better picture of the metabolic system's goals.

7.3 Results

7.3.1 Diauxic Growth Objective Function Analysis

This methodology is first tested by using it to determine the metabolic objective function for a well-known model system - *E. coli* cultured in mixed substrate environments. Comparing the objective functions generated by this method with those which have been traditionally used to model *E. coli* such as growth rate maximization provides a rational method to validate this objective function identification method.

For these systems which are visualized in figure 7.3a and 7.3b, lumped cybernetic models were used to calculate hypothetical u_i variables for enzymes related to the consumption of competing substrates. These u_i variables were then compared with normalized gene expression trends. From this comparison a set of optimal weights, $\mathbf{w}_{opt}$, was generated by minimizing the SSE between u_i and the mean behavior of gene expression profiles for genes that are related to the uptake of different substrates.

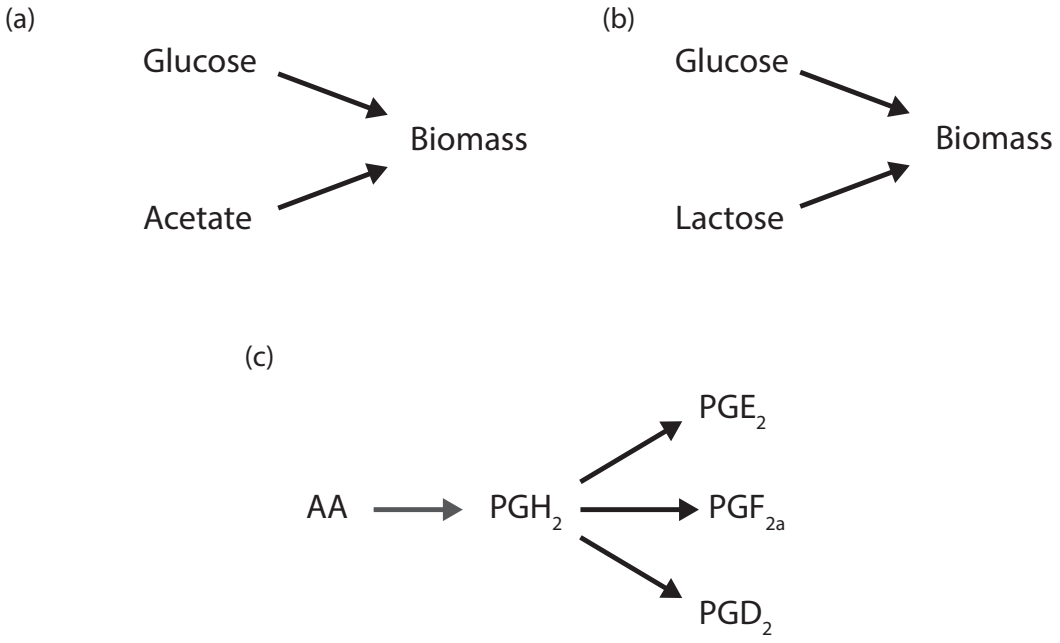

Figure 7.3.: Metabolic networks for reaction systems analyzed by the objective function search method.

Table 7.1.: Values of optimal weights from fitting u_i variables to gene expression data.

	Glucose	Secondary Substrate	
Glucose-Acetate	1	0.166	
Glucose-Lactose	0.1215	1	
	PGE2	PGF2	PGD2
Prostaglandin	1	1E-3	5.59E-3

7.3.1.1 Glucose-Acetate System

The first system of focus is *E. coli* growing in a mixed culture of glucose and acetate for which gene expression and growth data were available [65]. To calculate the u_i variables, substrate consumption rates were approximated using the initial substrate levels and changes in biomass. Via comparison of the u_i variables and relevant gene expression data, optimal weights were generated for this system and are identified in Table I. The weight for glucose consumption was higher in value compared to acetate which indicates that an objective weight of 1 to 1 for both pathways predicts an earlier transition in u_i variables than gene expression does. The improvement in fit of the control variables to gene expression is illustrated in figures 4a and 4b.

With these weights, a metabolic signal that is representative of the processes' changing goals, $\dot{J}$, is estimated. In order to determine the biological relevance of $\dot{J}$, it is compared to the changes in gene expression for various genes, $\dot{g}_k$. Pearson's correlation was calculated for all genes. From this, genes with statistically significant and positive values ($\rho > 0$ and $p < 0.05$) were separated into a gene list for further analysis. These genes were determined to be the most similar in changing expression compared to the changing goal. This gene list, found in the appendices, consists of 413 genes total and is visualized with the goal signal in figure 7.4b.

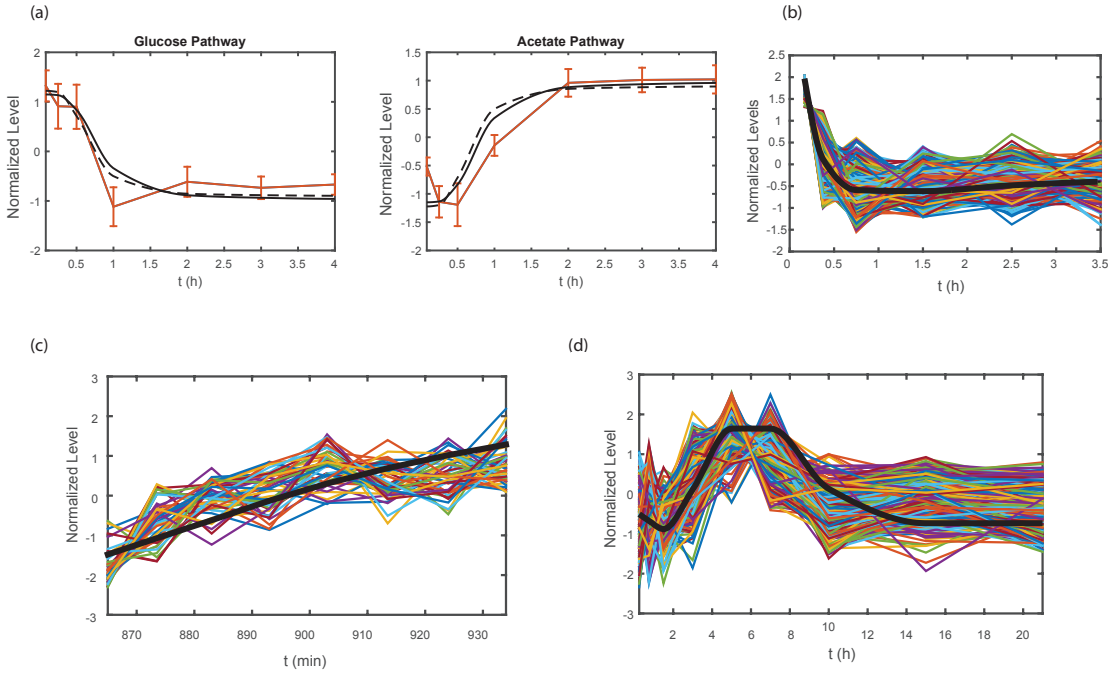

Figure 7.4.: Return on investment optimization and goal signals with gene sets. (a) Improvement of control prediction for glucose-acetate system where dashed lines (- -) represent equal weightings to pathway ROI and solid lines (-) represent optimal weighting for fit. In remaining subfigures, goal signal for each system is in black with colored lines for genes in the goal signal gene set for (b) glucose-acetate system (c) glucose-lactose system and (d) prostaglandin system.

Pathway enrichment analysis was performed using this list to establish which KEGG pathways were overrepresented. The top pathways that were enriched were: ABC transporters, TCA cycle, glyoxylate and purine biosynthesis. These enriched pathways are relevant to acetate consumption and previously used objective functions such as growth rate maximization. Further discourse of these enriched pathways will air in the subsequent discussion. A full list of enriched KEGG pathways and their enrichment statistics are available in Table 7.3.

7.3.1.2 Glucose-Lactose System

Analysis of the metabolic changes in the glucose-lactose system was made in a similar fashion to the glucose-acetate system for gene expression data [66]. In this scenario, substrate data was provided and u_i variables were directly inferred from the changing substrate data. Note that the regulated rates for each pathway were approximated by the negative rates of substrate uptake for each pathway. The optimized weights in this case showed the opposite behavior of the glucose acetate system with lactose having a higher weight than glucose. This indicates that the best fit of gene expression occurred when the objective function was weighted towards the lactose pathway for this particular data set. This is further explained in the subsequent discussion section.

The behavior of $\dot{J}$ and the genes that correlate to it are in figure 7.4c. This set of genes that correlates with the goal signal is stated in the appendices and consists of 46 genes. Enrichment analysis indicated that amino acid biosynthetic pathways and pyrimidine biosynthesis were the enriched pathways. Refer to table 7.3 for more information on the statistical nature of this relationship.

7.3.2 Prostaglandin Biosynthesis

Beyond well-characterized, bacterial systems, objective functions for more complex organisms are more difficult to hypothesize. In light of this, this method is ap-

Table 7.2.: Enriched pathways for glucose-acetate system

KEGG Pathway	Gene Count	%	P-Value	Benjamini
ABC transporters	22	5.6	1.3E-39	1.80E-36
TCA Cycle	11	2.8	3.50E-24	2.702E-22
Purine Metabolism	13	3.3	6.80E-24	5.00E-22
Two-component System	12	3	1.30E-19	5.70E-18
Glyoxylate and Dicarboxylate Metabolism	8	2	5.1E-16	2.20E-14
Pyruvate Metabolism	8	2	3.80E-15	1.80E-13

Table 7.3.: Enriched pathways for glucose-lactose system

KEGG Pathway	Gene Count	%	P-Value	Benjamini
Glycine, Serine and Threonine Metabolism	3	7.5	1.20E-3	0.27
Lysine Biosynthesis	2	5	2.3E-2	0.72
Alanine, Aspartate and Glutamine Biosynth.	2	5	4.49E-2	0.79
Pyrimidine Metabolism	2	5	8.4E-2	0.86

Table 7.4.: Enriched pathways for PG system

KEGG Pathway	Gene Count	%	P-Value	Benjamini
Endocytosis	7	5.7	4.10E-3	0.23
SNARE Interactions in Vesicular Transport	3	2.4	3.40E-2	0.66
Cytokine-Cytokine Receptor Interaction	6	4.9	3.80E-2	0.55
Notch Signaling Pathway	3	2.4	5.60E-2	0.6

plied to deductively determine the objective function of prostaglandin (PG) metabolism in RAW 246.7 macrophages, a cell line derived from *Mus musculus*. In this system, one precursor, arachidonic acid (AA) is converted into downstream prostaglandins PGD_2, PGF_{2a}, and PGE_2 in scenarios where an infectious stimulus present [83]. This infectious stimulus is Kdo_2-lipid-A (KLA) which is similar to the molecules that cover the surfaces of pathogenic bacteria. When KLA binds to a macrophage's Toll-like receptor 4 (TLR4), the macrophage responds to fight the perceived infection. While one may be able to guess at general objective functions for this situation such as maximize the rate of inflammation generating signals, it is difficult to implement this in a precise way.

In this system, one substrate is converted into three products which is visualized in figure 7.3c. The optimal weights were generated directly from the metabolite data using the product formation rate and previously fit decay parameters to designate the regulated rate for each metabolic pathway. The gene expression for each pathway's enzymes were determined by identifying key genes for metabolic enzymes using the KEGG pathway database. The optimal weights are in Table I. These weights show that the system has a strong regulatory weighting towards the pathway that generates PGE_2.

Using the optimal weights, the metabolic goal signal from the network was approximated and tested for correlation with all genes to generate a gene set. This is visualized in figure 7.4d. The analysis of this gene set yielded the following explanations for metabolic goal: endocytosis, SNARE interactions in vesicular transport, cytokine-cytokine receptor interaction, and the notch signaling pathway. The statistical significance and FDR are stated in table 7.4.

7.4 Discussion

7.4.1 Glucose-Acetate Diauxie

The explanations provided by pathway enrichment analysis for the change in metabolic goal signal $\dot{J}$ in the glucose-acetate diauxic growth scenario interestingly revolve around pathways that are upregulated during *E. coli*'s transition from growth on glucose to growth on acetate. Flux analysis of the changes in metabolism show that TCA cycle and glyoxylate are upregulated during this transition [70]. The fact that the metabolic goal indicated by $\dot{J}$ during this transition reflects the significant metabolic changes that occur helps to validate the idea that pathway analysis is useful in making sense of the weighting scheme for this system.

Furthermore, the enrichment of purine metabolism and ABC transporters pathways helps to validate the idea that well-established objectives like growth rate maximization [3, 33] and substrate uptake maximization [19, 23]. The synthesis of nucleotides enables growth via DNA and protein synthesis via mRNA, tRNA and rRNA and is a critical component of the survival of genes. RNA and DNA make up a significant portion of cell's macromolecular compositions and ribosomes, composed of rRNA, are a major component of growth [30]. The uptake of substrates through ABC transporters indicates a possible goal of promoting the scavenging of other sources of nutrients [65] which is reflective of hypothesized goals related to resource competition.

7.4.2 Glucose-Lactose Diauxie

Pathway enrichment analysis of the glucose-lactose diauxic system yielded less statistically significant findings. This likely stems from two factors a smaller that desired gene list (<100) and the behavior of the gene expression data. It is possible that these two factors are related. In reference to the behavior of the data, the microarray data has numerous kinks in the time course going up and down and up and down which makes the gene behavior less likely to significantly correlate with the

smoother profile for u_i. Other time-series expression data, such as that used in the other two test cases, does not show this kinked behavior.

Despite the lower significance in statistical data, the pathways that show similarity to the goal signal are those for amino acid synthesis and pyrimidine biosynthesis. The presence of pyrimidine biosynthesis in the set of enriched pathways helps to confirm what is observed in the glucose-acetate system - an enrichment of nucleotide synthesis which, as stated previously, is a key component of growth. This helps to confirm the idea that more general objective functions like growth rate maximization are relevant. The generation of amino acids is consistent with the notion of biomass maximization as proteins constitute a majority of the macromolecules that make up cells [114].

One important item of note is the fact that the metabolic goal signal weighted different pathways inconsistently between the different scenarios tested. In the glucose-acetate system, glucose had the highest weighting while lactose had the highest weighting in the glucose-lactose system. This difference is due to the fact that the timing of gene expression was different with respect to the timing of glucose depletion. The u_i variables were delayed in their transition times for a better fit in the glucose-acetate case. It also means that the u_i variables for the glucose-lactose best fit the data when they described an earlier transition between gene expression for the two pathways. To accommodate the fits of the gene expression data, the goal-signal profile shifts between the two systems. On the surface, this discrepancy should not exist but it is conceivable that the metabolic changes during the shifts between these two separate substrates are better explained by different objective functions. Other work has shown that there is no single yield-based objective function that is universally accurate for a range of systems and conditions [48].

There may be several sources of error that contribute to this difference in goal signals. It may be due to differences in how the substrate trends were extrapolated from the data. If the substrates showed different timing in their depletion, the goal signal could have shifted to accommodate this timing difference. Another possible explanation is the fact that glucose genes do not show a strongly similar relationship

to the u_i variable in this diauxic growth scenario. The metabolic overlap in pathways for the consumption of glucose and lactose is significant with only the *lac* operon and the glucose transporter ptsG being outside of this overlap. This could make it harder for a goal signal to be generated by looking strictly at the gene expression profiles for this system.

7.4.3 Prostaglandin Metabolism

In the context of a macrophage reacting to infection, the enrichment of the endocytosis, SNARE interactions in vesicular transport, cytokine-cytokine receptor interaction pathways have meaning. Endocytosis and SNARE interactions [115] fit into the macrophage response to infection in that macrophages fold their cellular membranes to envelop invasive bacteria. This is further confirmed when relaxing the filtering criteria for the gene list from a correlational significance (p-value =0.10 from p-value =0.05). With the new gene list, the pathway for Fc Gamma R-mediated Phagocytosis is enriched (p-value 1.9e-02, Benjamini=3.5e-01). A full list of these pathways enriched for the enriched pathways for the gene list provided by this less stringent filtering criterion is found in Table 7.5.

The other striking enriched pathway of the set is for the cytokine-cytokine receptor interaction pathway. This makes sense as macrophages produce and sense a variety of cytokines to cope with infection. This is also confirmed by relaxing the same filtering criteria extend the gene list for the metabolic goal. When this happens, chemokine signaling pathway is also enriched which embodies the downstream interactions of the cytokine-cytokine receptor interaction pathway. Additionally, the toll-like receptor-signaling (TLR) pathway is enriched in the extended list which is directly related to the condition of interest [87].

The relevance of the enriched pathways, determined by each transcript's relationship to the goal profile, to the macrophage's well-established response to KLA help to validate the idea of a metabolic goal signal. Furthermore, there is a large body of liter-

Table 7.5.: Results of pathway enrichment analysis for gene list generated using relaxed filtering criterion $p - value < 0.10$.

KEGG Pathway	Gene Count	%	P-Value	Benjamini
Chemokine Signaling Pathway	10	3.4	2.20E-3	0.18
Endocytosis	10	3.4	4.50E-3	0.18
Gap Junction	6	2	1.10E-2	0.290
Fc gamma R-mediated Phagocytosis	6	2	1.9E-2	3.5
Snare Interactions in Vesicular Trans.	4	1.4	2.20E-2	3.20E-3
Notch Signaling Pathway	4	1.4	4.40E-2	0.43

ature that links prostaglandin metabolism to the pathways for phagocytosis [116–119], chemokine signaling [120–123] and TLR 4 [87, 124].

7.5 Conclusions

The fact that goal signals have significant similarity with what is hypothesized about the functions of various metabolic networks makes a strong case for using them to describe the regulation of metabolic systems. Also, given that these weighting schemes accurately reflect the changes in gene expression for the pathways of focus and can be used to generate goal signals that can be matched to highly relevant metabolic pathways, they should allow for a more complete cybernetic description of biological regulation. Analysis of gene lists related to the goal signal in all three scenarios yielded feasible biological explanations the gene lists that matched the goal signal. Both the glucose-acetate and glucose-lactose systems showed metabolic goals related to DNA synthesis, a key component of cellular growth. Other enriched pathways were explained by information from the literature.

This work defines a novel method to deductively determine dynamic cybernetic objective functions. Furthermore, it provides a rational procedure for identifying the biological significance of the deduced objective. The ability to identify objective functions for metabolic systems where the metabolic goal is hard to hypothesize could facilitate the development of cybernetic models for complex metabolic systems such as those present in mammalian cells. This approach has potential applications ranging from bioprocessing to pharmacology.

8. GOING BEYOND NONLINEAR OPTIMIZATION FOR PARAMETER ESTIMATION USING LINEARIZED REACTION KINETICS

Summary

To address some of the issues related to parameterization that have arisen in chapters 5 and 6, this chapter attempts to provide a feasible solution for the parameter estimation problem inherent to Hybrid Cybernetic Models. Instead of relying on cumbersome nonlinear optimization methods to minimize some error function, the following uses an analytical approach to break down cybernetic model kinetics into a linearized system of equations that yields a single linear least-squares solution for parameter estimation. This contrasts with foregoing methods of parameterization for cybernetic models (i.e. pattern searches and gradient searches) in that the solution is computationally trivial and yields only one estimate of parameter values.

8.1 Introduction

Cybernetic models are distinguished from other methods that describe metabolism in that they include an approximation metabolic regulation via the formulation dynamic metabolic goals that guide organisms to maximize some product of metabolism over time. When models of this nature were first proposed to describe the diauxic growth of bacteria on mixtures of multiple substrates, their parameterization relied upon the characterization of growth on individual substrates [3]. As cybernetic models have advanced in complexity to accommodate large networks of intracellular reactions using elementary flux modes (EFMs) [18], the parameterization of these models has become less straightforward as multiple EFMs compete for the same substrate and no experiment can characterize the uptake of substrate into a single EFM. Given this challenge, the specification of Hybrid Cybernetic Models (HCMs) [19, 125] and

Lumped Hybrid Cybernetic Models (L-HCMs) [23,24] require nonlinear optimization to estimate the kinetics of uptake into different EFMs. Identifying model parameters in this fashion can be computationally cumbersome and minimization searches do not guarantee that the search will conclude at a global minima of the search function.

In the identification of parameters for kinetic models of metabolism, this same issue occurs the where the parameter search function has multiple local minima. Validation of parameters relies on predicting metabolic states on which the model was not trained [126]. A variety of approaches to determine parameters for kinetic models have been proposed such as curated databases [127] ensemble modeling [32, 128], Bayesian Parameter Estimation [129], and ^{13}C labeling experiments during nonstationary states [130] among many others. Each of these methods come with their respective pros and cons and are not readily applicable for application to HCM or L-HCM in that the dynamic uptake of substrate into various EFMs is an abstraction of the biophysical process. In other words, it is difficult to identify explicit facets of a biological process that are parallel to EFMs on which one can directly characterize parameters.

While other dynamic models of metabolism have had solutions proposed for the identification of parameters, there still exists a need for better parameter specification for cybernetic models. To remedy this, the present work proposes a method that works for cases where the kinetics can be transformed into a linear function of the analyzed and transformed. This can be readily applied to the Michaelis-Menten kinetics that are typically used to describe the uptake of substrates into different EFMs [19,23,24] and extended to other uptake kinetic expressions. By specifying the following parameter identification approach, the computational burden of specifying cybernetic models is reduced and the confidence in the resulting parameters values is increased.

8.2 Methods

8.2.1 Determination of Unregulated Uptake Fluxes

A cybernetic model relies on time-course data for its identification. The ultimate aim of the method is to determine the unregulated uptake fluxes (UUFs), $\mathbf{r}_{\mathbf{M}}^{kin}$, from data for the time-profiles of the changes in substrates, products and cell concentrations. These UUFs are a kinetic expression that relates the amount of substrate, s_i, to its uptake through the i^{th} EFM through some expression such as a Michaelis-Menten one like

$$r_{M,i}^{kin} = e_i \frac{k_i s_i}{K_i + s_i}, \tag{8.1}$$

where e_i is the bulk enzyme for catalysts for the i^{th} enzyme and k_i and K_i are parameters that we would like to establish. In establishing the UUFs from the data, we can subsequently manipulate them to solve for the kinetic parameters in a robust fashion.

These concentrations of substrates, products and cell concentrations are defined in some vector $\mathbf{x}$ of m species as

$$\mathbf{x} = \begin{bmatrix} \mathbf{s} \\ \mathbf{p} \\ c \end{bmatrix} \tag{8.2}$$

where $\mathbf{s}$ is the vector of substrates, $\mathbf{p}$ is a vector of products and c is the concentration of cells. To determine the UUFs, it is necessary to manipulate the model using data for $\mathbf{x}$.

In HCM and L-HCM, the changes in $\mathbf{x}$ and the regulated uptake fluxes are related through the coupled system of differential equations

$$\frac{1}{c}\frac{d\mathbf{x}}{dt} = \mathbf{S}_{\mathbf{x}}\mathbf{Z}\mathbf{r}_{\mathbf{M}}, \tag{8.3}$$

in which $\mathbf{S}_{\mathbf{x}}$ is the stoichiometric matrix that describes only species in $\mathbf{x}$, $\mathbf{Z}$ is the matrix describing n EFMs or lumped EFMs, in the case of L-HCM, and $\mathbf{r}_{\mathbf{M}}$ is the set of n regulated uptake fluxes through the EFMs. Given complete knowledge of $\mathbf{x}$ and

an ability to extrapolate $\frac{d\mathbf{x}}{dt}$ from $\mathbf{x}$, it is possible to characterize the regulated uptake rates $\mathbf{r_M}$ for each EFM given that the number of EFMs is equal to the number of species in $\mathbf{x}$ or

$$m = n. \tag{8.4}$$

This will make the product $\mathbf{S_x Z}$ invertible. To enforce the above equality, one can select a subset of EFMs that span the maximal volume of the convex hull that makes up the yield space for the EFM set for HCM using the same approach as is specified in yield analysis [22] or distinguish a lumping scheme that determines m lumped EFMs for L-HCM. Given that the number of EFMs is equal to the number of species in $\mathbf{x}$, equation 8.3 can be manipulated for the regulated uptake fluxes with

$$\mathbf{r_M} = \frac{1}{c}(\mathbf{S_x Z})^{-1}\frac{d\mathbf{x}}{dt}. \tag{8.5}$$

The regulated uptake fluxes are related to the unregulated uptake fluxes by the following relationship:

$$\mathbf{r_M} = \mathbf{V}\mathbf{r}_\mathbf{M}^{kin}, \tag{8.6}$$

where $\mathbf{V}$is an n by n diagonal matrix for the n v_i cybernetic variables that describe enzyme activity regulation. This cybernetic variable, called the proportional law, takes the form

$$v_i = \frac{p_i}{\max_j p_j}, \tag{8.7}$$

where p_i represents the return on investment (ROI) for each pathway. ROI can take a variety of forms such as maximization of growth rate

$$p_i = Z_{i,G} r_i^{kin}, \tag{8.8}$$

where $Z_{i,G}$ represents the amount of biomass or growth created by the i^{th} elementary model. Other forms of ROI are also feasible which is generalized to

$$p_i = w_i r_{M,i}^{kin}, \tag{8.9}$$

in which w_i represents some weight related to the metabolic goal.

For each EFM uptake, the following relationship exists:

$$r_{M,i} = v_i r_{M,i}^{kin} \tag{8.10}$$

This can be manipulated to determine the regulated uptake fluxes through

$$r_{M,i} = v_i r_{M,i}^{kin} \tag{8.11}$$

$$r_{M,i} = \frac{p_i}{\max_j p_j} r_{M,i}^{kin} \tag{8.12}$$

$$r_{M,i} = \frac{w_i (r_{M,i}^{kin})^2}{\max_j w_j r_{M,j}^{kin}} \tag{8.13}$$

$$r_{M,i}^{kin} = \sqrt{\frac{r_{M,i} \max_j w_j r_{M,j}^{kin}}{w_i}} \tag{8.14}$$

From this, each EFM's UUF can be established in that w_i is postulated by the modeler as an assumption for a feasible objective function and $\max_j w_j r_{M,j}^{kin}$ is known from the property that the maximum ROI can is known directly from the data. This is because

$$\arg\max_i (w_i r_{M,i}) = \arg\max_j (w_j r_{M,j}^{kin}), \tag{8.15}$$

given that the pathway with the maximal ROI will have no downregulation of the unregulated uptake flux rate or $v_i = 1$. Note that because the UUFs can be determined at each point in time, the ROIs for each EFM can also be determined at each point in time.

8.2.2 Extracting Enzyme Profiles from Data

In HCM and L-HCM, the changes in enzyme level are established by

$$\frac{de_i}{dt} = \alpha_i + u_i \frac{k_{E,i} s_i}{K_{E,i} + s_i} - (\mu + \beta_i) e_i \tag{8.16}$$

where α_i is the rate of constitutive enzyme synthesis, and the last term represents depletion of enzyme due to growth μ and degradation β_i. In this expression, the parameters α_i and β_i can be taken from prior work [19]. Growth rate, μ can be extrapolated from data for time-dependent changes in cell concentration c as

$$\mu = \frac{d \ln c}{dt}. \tag{8.17}$$

The second term in enzyme equation 8.16 represents the controlled rate of induced enzyme synthesis and is modulated by cybernetic variable u_i which is a relationship between the ROIs

$$u_i = \frac{p_i}{\sum_j p_j}, \tag{8.18}$$

and means that the fractional allocation of resources for the synthesis of the i^{th} pathway's bulk enzyme is proportional to the fractional ROI endowed by investment into that pathway. Given that p_i is known for each pathway at each time, this cybernetic variable is also known for each time point in the data. The Michaelis-Menten expression for induced enzyme synthesis is also known given the assumption that parameters $k_{E,i}$ and $K_{E,i}$ are the same from previous applications of HCM and L-HCM. The substrate concentration is also known at each time. Therefore, the entire expression for $\frac{de_i}{dt}$ is characterized at each time point given some starting guess for each $e_{i,0}$. From this, the enzyme concentrations can be numerically approximated using numeric integration for each pathway's enzyme quantity.

8.2.3 Determination of Rate Parameters

Given that the unregulated uptake fluxes, substrate level and enzyme quantities are known for each point in time, it is possible to extract the kinetic parameters by linearizing the kinetic expression by taking its inverse or

$$\frac{1}{r_{M,i}^{kin}} = \frac{K_i}{k_i}\frac{1}{e_i s_i} + \frac{1}{k_i}\frac{1}{e_i}. \tag{8.19}$$

The behavior of the set of observed fluxes can be generalized into a linear system for each observed time point t_s as

$$\mathbf{y} = \mathbf{X}\mathbf{b} \tag{8.20}$$

or

$$\begin{bmatrix} \frac{1}{r_{M,i}^{kin}}(t_1) \\ \frac{1}{r_{M,i}^{kin}}(t_2) \\ \dots \\ \frac{1}{r_{M,i}^{kin}}(t_p) \end{bmatrix} = \begin{bmatrix} \frac{1}{e_i s_i}(t_1) & \frac{1}{e_i}(t_1) \\ \frac{1}{e_i s_i}(t_2) & \frac{1}{e_i}(t_2) \\ \dots & \dots \\ \frac{1}{e_i s_i}(t_p) & \frac{1}{e_i}(t_p) \end{bmatrix} \begin{bmatrix} \frac{K_i}{k_i} \\ \frac{1}{k_i} \end{bmatrix} \tag{8.21}$$

To solve for the parameters in $\mathbf{b}$, one can quickly find a least squares estimate using the Moore-Penrose pseudoinverse for $\mathbf{X}$

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{8.22}$$

From $\mathbf{b}$, we can establish each EFM's uptake flux kinetic parameters k_i and K_i where

$$k_i = {b_2}^{-1} \tag{8.23}$$

$$K_i = \frac{b_1}{b_2} \tag{8.24}$$

8.2.4 Linearization of Other Kinetic Expressions

This method can be used to fit parameters for other kinetic uptake expressions. For example, if an EFM relies on the uptake of two substrates, the kinetics could take some multiplicative form of the Michaelis-Menten kinetics such as

$$r_{M,i}^{kin} = e_i k_i \frac{s_1}{K_{1,i} + s_1} \frac{s_2}{K_{2,i} + s_2} \tag{8.25}$$

The inverse of this kinetic mechanism yields a more easily analyzed form of the kinetics

$$\frac{1}{r_{M,i}^{kin}} = \frac{K_{1,i}K_{2,i}}{k_i} \frac{1}{e_i s_1 s_2} + \frac{K_{1,i}}{k_i} \frac{1}{e_i s_1} + \frac{K_{2,i}}{k_i} \frac{1}{e_i s_2} + \frac{1}{k_i} \frac{1}{e_i} \tag{8.26}$$

which yields the same analysis as in equation 8.21 where the system of equations is

$$\begin{bmatrix} \frac{1}{r_{M,i}^{kin}}(t_1) \\ \frac{1}{r_{M,i}^{kin}}(t_2) \\ \dots \\ \frac{1}{r_{M,i}^{kin}}(t_p) \end{bmatrix} = \begin{bmatrix} \frac{1}{e_i s_1 s_2}(t_1) & \frac{1}{e_i s_1}(t_1) & \frac{1}{e_i s_2}(t_1) & \frac{1}{e_i}(t_1) \\ \frac{1}{e_i s_1 s_2}(t_2) & \frac{1}{e_i s_1}(t_2) & \frac{1}{e_i s_2}(t_2) & \frac{1}{e_i}(t_2) \\ \dots & \dots & \dots & \dots \\ \frac{1}{e_i s_1 s_2}(t_p) & \frac{1}{e_i s_1}(t_p) & \frac{1}{e_i s_2}(t_p) & \frac{1}{e_i}(t_p) \end{bmatrix} \begin{bmatrix} \frac{K_{1,i}K_{2,i}}{k_i} \\ \frac{K_{1,i}}{k_i} \\ \frac{K_{2,i}}{k_i} \\ \frac{1}{k_i} \end{bmatrix}. \tag{8.27}$$

The pseudoinverse can then be used to determine the parameter vector in this equation. From there, individual parameters can be identified through algebraic manipulation of the parameter vector.

Other kinetic formulations including various ways of modeling inhibition (e.g. competitive, noncompetitive, uncompetitive) can also be treated in a similar fashion.

8.3 Results

For proof of concept, this method was used to determine the parameters for a simplified version of the HCM model for *E. coli* GJT001 [19]. This simple model only tracks the changes in concentration of a subset of all external products. The selected subset consisted of the species glucose, biomass, acetate and ethanol. Acetate and ethanol were chosen because they embodied the most abundant products generated by the strain of bacteria. To allow for the inversion of the product of the external species' stoichiometric matrix and the EFM matrix, four elementary modes were selected from the set of 8. This selection procedure occured in a fashion similar to yield analysis. Towards this, the yield space of acetate on glucose and ethanol on glucose was analyzed and a subset of 4 EFMs that maximized the area of the convex hull was chosen. This selection procedure retained 96.82% of the 2-dimensional yield space's original area. This is visualized in figure 8.1.

The retained EFMs are in Table 8.1. From these 4 EFMs, the parameters for the unregulated rates were developed using the method from the previous section. The uptake kinetics for each EFM is composed of a simple Michaelis-Menten expression in the form of equation 8.1. Each mode's k_i and K_i are listed in table 8.2. The $\mathbf{S_x}$ matrix can be extrapolated from [19] taking the rows representative of the metabolites glucose, acetate, ethanol and biomass.

Because the inverse of the stoiciometric-EFM product yielded negative values for the regulated uptake rates, the routine *lsqnonneg* in MATLAB was employed to extrapolate the regulated rates from the changes in metabolite concentrations which forced the regulated rates to be positive in value. Following this, the enzyme levels

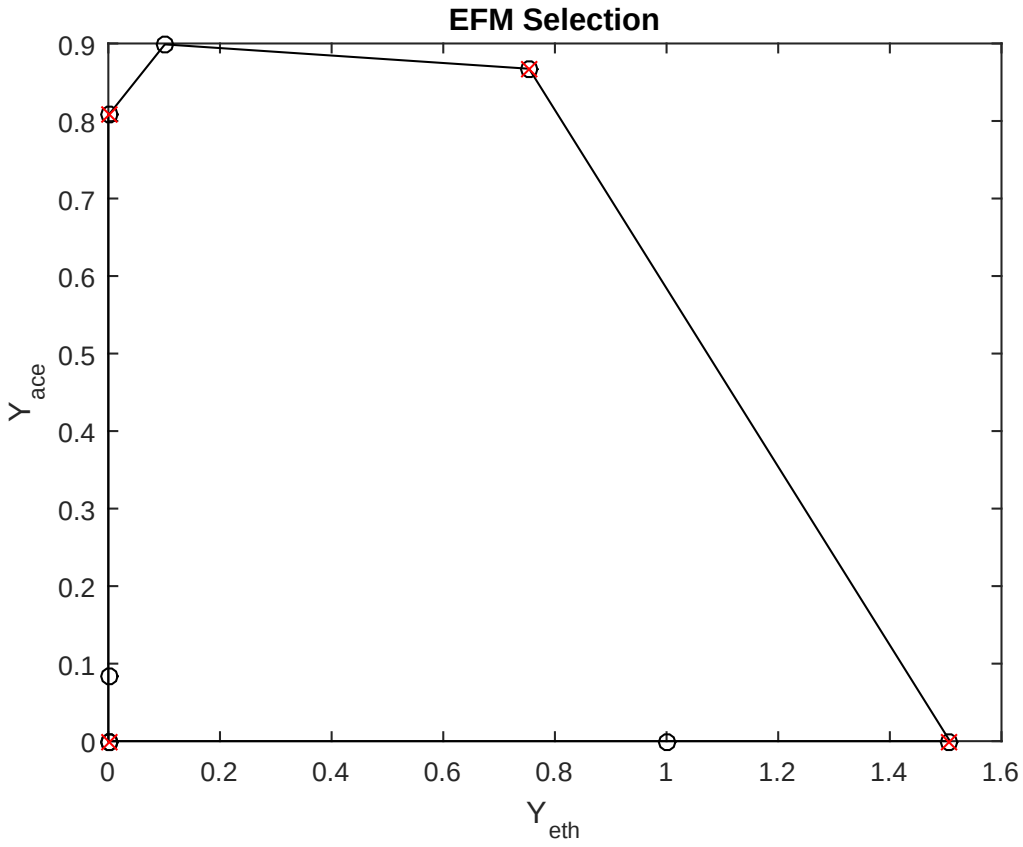

Figure 8.1.: Visualization of yield space analyzed to reduce the original set of 8 EFMs to 4. All EFMs are denoted with black circles and retained EFMs are represented by a red x.

Table 8.1.: Elementary Flux Modes used in model developed.

51.61	44.158	0.678	88.992	37.38	0	0	88.992	0	0	0	1
49.239	42.464	0	84.927	0	35.688	39.754	9.485	39.754	0	0	1
35.761	28.986	0	57.973	22.212	0	57.973	0	31.019	26.954	0	1
59.026	36.741	15.509	88.992	29.966	0	88.992	0	0	88.992	0	1

Table 8.2.: Model Parameters - Simplified HCM model for *E. coli* GJT001.

Mode	k_i	K_i
1	0.5481	99.8515
2	0.2241	42.9605
3	0.2943	0.01
4	0.1497	0.01

and parameters were determined. The parameter values are in table 8.2 and their fit is visualized in figure 8.2. It is evident in figure 8.2 that the fit achieved by this method describes the observed data well.

8.4 Discussion

Compared to nonlinear optimization approaches where parameters are solved by guessing parameters to minimize some error function in a generic least-squares format like

$$\arg\min_\theta \sum_s \left(\mathbf{y}_s - \hat{\mathbf{y}}_s(\theta)\right)^2, \tag{8.28}$$

in which $\mathbf{y}_s$ represents the data at time, t_s, and $\hat{\mathbf{y}}_s(\theta)$ represents the model prediction of the data at that same time. Gradient searches [131], genetic algorithms [132], pattern searches [133], and other methods all require large numbers of parameter guesses to find some minimum of the error function. Each of these parameter guesses, in turn, necessitates the simulation of the set of differential equations which can be often be stiff in the case of cybernetic models meaning that guessing thousands of parameters takes a non-trivial amount of time. Moreover, once the minimization algorithm establishes the parameters, it does not provide any insurance that a global minimum has been found.

Prior work, in this thesis, such as for the determination of the prostaglandin parameters, needed days of computation to simulate. The presented method of param-

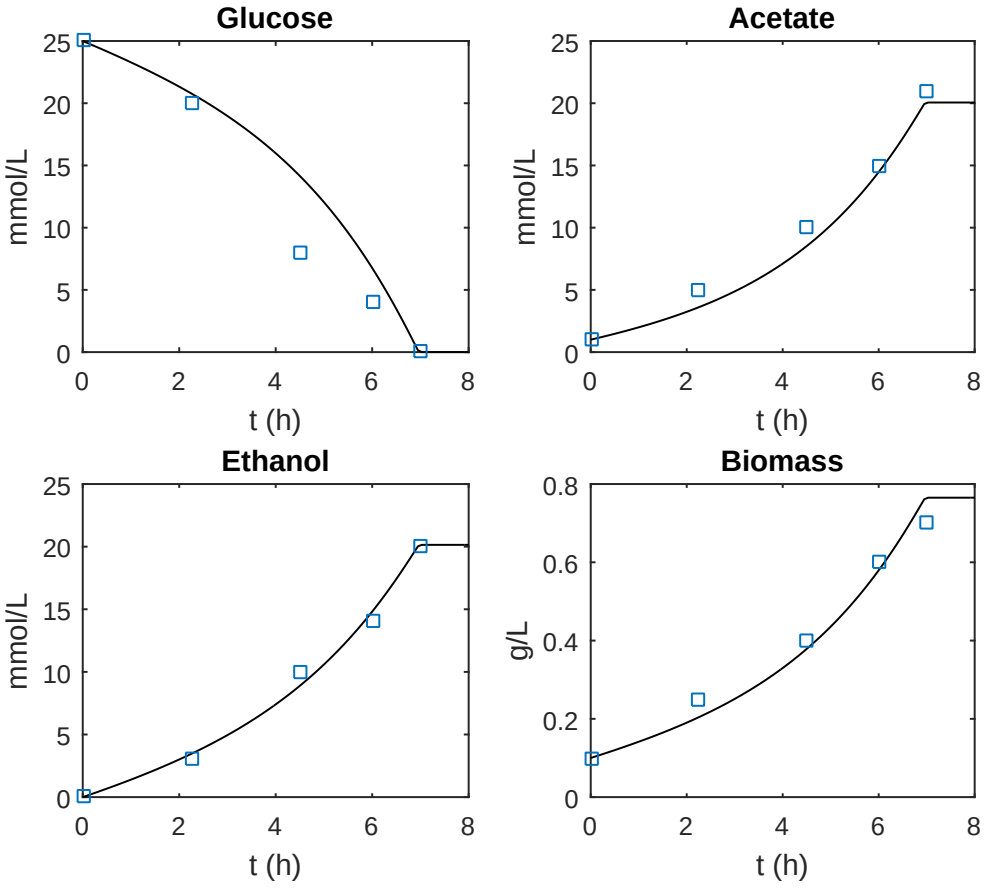

Figure 8.2.: Fit of model using parameters established from this method. Data points are marked by squares.

eterization needs the numeric integration of only one system of differential equations and the calculation of a pseudoinverse takes seconds to compute which contrasts with prior work which takes significantly longer for parameter determination. Moreover, there is only one solution in parameter space provided by this method given an input of some set of data, a hypothesized cybernetic objective function, a guess at initial enzyme levels and enzyme parameters used from prior work. The linear-least squares method also provides the global minimum for the observed data and model.

Having effective ways of specifying a cybernetic model's parameters is important to the cybernetic framework as a whole. Nonlinear optimization approaches are cumbersome, inefficient and do not endow a modeler with confidence of the absolute nature of parameter values. This method should be practical for any modeler seeking to parameterize a cybernetic model and allows for flexibility in the initial modeling assumptions. Instead of encountering the rigmarole of nonlinear optimization for each objective function assumption, one can easily develop models and test for the robustness of numerous objective functions.

Overall, this refined method of parameter estimation should enable modelers to more quickly and effectively develop descriptions of metabolic phenomena using the notion that regulation seeks dynamic optimality. It is clear from the fit achieved in figure 8.2 that the method works. Facile model identification allows for more complete probing of questions related to the truth of dynamic optimality in cells because less effort is expended on parameter estimation - a significant portion of the overall effort needed to generate cybernetic descriptions. Easier parameter identification could help make more conclusive statements about the fundamental nature of metabolic goals using analysis similar to what takes place in chapter 6.

9. SUMMARY

The goal of this dissertation was to establish whether cybernetic control provides a good description of biological regulation by probing research questions related to three main areas: the ability of cybernetic models to compress data relative to other models, the accuracy of the enzyme synthesis control variables in reference to data that is representative of enzyme synthesis control, and lastly, the ability to mine data for feasible objective functions. To phrase this as a question, when one considers the data generated by biological systems, is there evidence that metabolic changes are driven by goals? The analysis herein shows that there is definitive evidence that goals are a comprehensive way of modeling and understanding metabolism.

The first test of goals was in establishing that cybernetic models provide a compact description of biological processes relative to other models. The compression of data achieved by cybernetic models was higher relative to constraint-based and kinetic approaches. This shows that dynamic objective functions are more effective at capturing the regularity in data generated by metabolic systems. To reiterate, capturing the regularity in data generated by a process represents an understanding of the generating process. Models that capture more regularity (i.e. those that provide higher compression of the process's data) can be understood as having a deeper understanding of the generating process. Therefore, the analysis of relative compression of data by these different models makes the statement that metabolic goals yield a more complete picture of metabolic systems.

In the second segment of this work, a closer look at the cybernetic variable u_i was taken to determine if relevant biological processes for enzyme synthesis showed similar behavior. Two sets of systems were probed to establish this similarity. One set was composed of three studies of diauxic growth systems for different types of gene

expression data. This first study demonstrated the validity of cybernetic control of enzyme synthesis in this model system. To see if cybernetic control of enzyme synthesis was an accurate approximation of regulation in more complex mammalian systems, a similar study was made for prostaglandin metabolism in RAW 264.7 macrophages from *M. musculus*. This analysis also showed a statistically significant relationship between changes in the u_i variables and gene expression changes. Because of the observed likeness between cybernetic control variables for enzyme synthesis and gene expression data across multiple systems, it can be said that metabolic goals are a good description of metabolic control.

The last angle taken in addressing the question of cybernetic control's veracity was trying to deduce metabolic goals through the mining of large volumes of biological information. This method introduced the notion of "optimal returns on investment" and that of a generic metabolic "goal signal" which changes over time that can be compared with volumes of dynamic gene expression data. Applying these concepts to diauxic growth systems and prostaglandin metabolism generated feasible metabolic objective functions that were relevant in the context of the metabolic systems that were studied. The mining of data for objective functions is uniquely utile to the cybernetic framework in that all prior work in developing cybernetic descriptions of metabolism relied on a modeler's postulation of an appropriate objective function. Furthermore, the fact that relevant objective functions can be mined from data shows that biological systems indicate that there is regulation towards goals.

Overall, these approaches at validating cybernetic control demonstrate that there is reason to believe that the many iterations of biological refinement that have occured over countless generations have yielded optimal behavior. While it is impossible to answer the question "Are metabolic systems goal-directed?" with absolute certainty, this dissertation has shown that this concept is indeed true through the aforementioned lines of reasoning.

10. FUTURE WORK RECOMMENDATIONS: EXTENSIONS OF CYBERNETIC MODELING TO TRANSLATIONAL RESEARCH

Summary

Mathematic models that capture the dynamics of metabolism offer much promise in providing deeper insight into medical applications. While kinetic frameworks are hindered by nontrivial complexity, cybernetic models offer a simpler approach in describing metabolic changes by means of simplifying intricate regulatory phenomena into the assumption of a dynamic objective function. Cybernetic models still endow their users with an ability to make predictions of numerous metabolic phenomena despite this simplification.

In this dissertation, cybernetic models have demonstrated an ability to simultaneously track the changes in gene expression and metabolism in mammalian cells which demonstrates the potential for cybernetic models to forecast changes in complex, pharmacologically relevant systems. Because of this, it is proposed that cybernetic models be applied to the investigation of drug-metabolism interactions.

The product of this effort should yield predictions of metabolic behavior during perturbations to metabolic systems with drugs. These projections of metabolic behavior include changes in the formation of metabolic products that are indicative of drug side-effects. The identification of models that describe metabolic perturbations due to drugs should also generate crucial understanding of how drugs affect cellular regulation. Classifying the nature of cellular regulation during drug-induced metabolic dysfunction allows for a more comprehensive approach to pinpointing drug targets, treatment timing, and dosage.

10.1 Introduction

Metabolism is a vastly complex process that is fundamental to life. Having a clearer picture of metabolic behavior enables a variety of societally important technologies ranging from pharmaceutical applications to biofuels. Our understanding of the intricate machinery that drives metabolic processes is accelerating in rate due to the advent of numerous omic technologies that allow researchers to get increasingly detailed pictures of biological processes. There is a need for models that give new insight into the volumes of biological data that are being generated by these technologies [134].

Cybernetic models of bacterial metabolism, armed with the assumption that metabolic pathways are regulated towards maximizing the rate of growth or substrate uptake, have yielded numerous predictions of intriguing metabolic phenomena. These models have provided for the predictions of complex substrate uptake patterns [3,9], the prediction of multiple steady states [14,21], metabolic fluxes [26], and, in chapters 4 and 5, the prediction of gene expression phenomena in both bacterial and mammalian systems.

These recommendations of future work seek to further develop cybernetic models for medical applications. Cybernetic models are just beginning to be approximate the behavior of mammalian systems which are germane to clinical applications. Recent approaches have been developed to learn objective functions from omic data and can be readily applied to these systems. The overall goal is to use cybernetic models to robustly predict how drugs perturb metabolic behavior in mammalian systems.

10.1.1 Significance

Systems biology has a number of promising applications that could help to transform the practice of medicine. Metabolic models can empower researchers to make sense of biological data and hypothesize connections between biomedical research and patient outcomes [135]. Tangent to approximation of metabolic systems, a number

of systems approaches have studied how diseases affect the behavior of regulatory networks [136]. The combination of metabolism and regulation are highly relevant to human health [137]. Cybernetic models offer an effective way to approximate the dynamics of both metabolites and regulation of metabolic reactions.

Other constraint-based metabolic modeling have endowed translational researchers the ability to predict drug targets to fight cancer [138,139]. These approaches do not allow for researchers to gain a deeper understanding of how time-dependent changes in metabolites affect disease progression. In light of this, cybernetic metabolic models offer their users with the possibility of new discoveries.

10.1.2 Innovation

The proposed work will attempt to use cybernetic models to gain a more significant understanding of the way that drugs affect metabolism. Dynamic modeling of metabolism endows researchers with the ability to make insights into biological behavior and make predictions to guide experimentation. Mathematical descriptions of metabolic dynamics using detailed kinetic formulations have a number of significant restrictions including difficult parameterization, and computational complexity. For example, whole-cell models, capable of describing the time-dependent expression of entire genomes in simple cells such as *M. genitalium* [140], rely on parameter sets numbering in the thousands and still require the simplification of metabolic processes using yield-based objective functions.

Instead of exhaustively enumerating the multitude of features that modulate metabolism, cybernetic metabolic models strike a unique balance between model complexity and accuracy by modeling the dynamic regulation of metabolism using the assumption that such processes are goal-directed [45]. Cybernetic models are in a unique position to provide medically significant predictions. Recent work describing prostaglandin metabolism in a *M. musculus* macrophage cell line have yielded predictions of how combinations of infection and statins affect the formation of inflammatory lipids which has been enabled by advanced technologies that characterize

lipids [87,89]. Beyond this, the cybernetic model for this system is capable of simultaneously predicting relative changes in gene expression. Given that cybernetic models allow for the prediction of metabolic phenomena on multiple levels, it is conceivable that they can also provide for the prediction of how inhibitory drugs interact with metabolic systems.

10.2 Recommended Research Goals and Plan

Provided the preceding discussion into the merits of cybernetic modeling, attempts to verify the following hypothesis are recommended:

10.2.1 Hypothesis:

Cybernetic models allow for the robust prediction of the interactions of drugs that affect metabolism because they are capable of robustly forecasting the behavior of gene expression and metabolites during perturbations to metabolism.

10.2.2 Goals:

To validate the stated hypothesis, the following research goals are recommended:

- Goal 1: Make predictions of metabolic behavior during perturbations to metabolic systems with drugs. Predictions of metabolic behavior include changes in the formation of metabolic products that are indicative of drug side-effects.
- Goal 2: Identify the nature of cellular regulation during drug-induced metabolic dysfunction.

10.2.3 Recommended Research Plan

To address the first research goal, prostaglandin metabolism will be of focus. Prostaglandins (PGs) are generated during inflammation and bind to a number of receptors that cause a variety of responses in immune system cells through second

messengers [81]. This system is of significant medicinal relevance in that it is targeted by nonsteroidal anti-inflammatory drugs (NSAIDs) to inhibit cyclooxygenase (COX) enzymes and prevent the formation of PGs. Over 30 billion OTC doses of NSAIDs are sold in the US every year [82]. Formulation of a cybernetic model describing the effects of different drugs that perturb PG metabolism would have clinical utility in that it could more robustly predict the behavior of PG metabolism than other drugs. Beyond the specific application of describing the effects of NSAIDs, this exercise would provide proof-of-concept for the cybernetic approach for modeling drug-metabolism interactions. The provision of a modeling framework that robustly predicts metabolic regulation with a low degree of model complexity would empower the process of drug development.

As covered in chapter 5, an already existing cybernetic model that describes prostaglandin formation in RAW 264.7 macrophage cells is capable of simultaneously predicting gene expression and metabolic behavior. This robust approximation of PG metabolism can also be used to predict the behavior of PG metabolism in situations where cyclooxygenase is inhibited. A sample of what the model projects is stated in chapter 5 (figure 5.4). This finding has yet to be experimentally verified. Beyond this, the upstream effects of blocking PG formation could be modeled using cybernetics to complement the understanding side-effects of NSAIDs. PGs are just a facet of a larger network of signaling lipids called eicosanoids. If there is some unique accumulation of lipids in a larger model that tracks the formation of other eicosanoid products, a cybernetic model would be able to predict such a phenomenon. Understanding the accumulation of other products provides information on what side effects will occur. A dynamic model will also enable researchers to gauge the time-dependent formation of side effects.

Beyond providing predictions of the effects of drugs, modeling the effects of drugs on cellular regulation and metabolic behavior will provide deeper insight into how drugs perturb the processes that control metabolism. Cybernetic models are formulated from the perspective that cells modulate metabolism to achieve goals. These

goals are relevant to the organism's survival and have been finely tuned through the eons of evolution where natural selection picks the metabolic algorithms that achieve survival and propagation. When drugs perturb metabolic systems, they are interfering with the preprogrammed goal that drives metabolic regulation.

Regulation witnessed during drug-induced metabolic dysfunction may provide a novel perspective on metabolic control. There a number of interesting scenarios such as (a) the cells regulate as if there is no drug perturbation, (b) cells react to perturbation and still control metabolism towards the same goal or (c) some new metabolic objective function becomes apparent. Cybernetic models allow for the performance of this type of analysis in that they can be used explicitly to model these scenarios. Comparison of a cybernetic model's prediction of optimal control during drug treatment with data will reveal how rigid metabolic regulation mechanisms are.

Understanding, how regulation occurs during treatment with drugs supports the development of treatment modalities that are more effective. If metabolic regulation circumvents the full realization of a drug's therapeutic potential, cybernetic models are capable of forecasting this. If a drug is not fully functional due to metabolic regulation, a cybernetic model could diagnose combinations of drug treatment that are.

The following activities are recommended:

Goal 1:

a. Small Network Model

   (a) Experimental characterization of PG dynamics during treatment with NSAIDs. Metabolite and gene expression data for the following conditions need to be collected: a control, inflammatory stimulus and combination of inflammatory stimulus and NSAID. Variation in doses of NSAID are needed for model cross validation.

   (b) Development of cybernetic model to describe the changes in PG formation for the different conditions.

(c) Validation of the model to test to see if it can predict PG changes for different NSAID dosing.

b. Large Network Model

(a) Characterization of a larger set of eicosanoids in the same set of conditions.

(b) Formulation of cybernetic model for this system.

(c) Cross-validation study as in item above.

(d) Analysis of perturbed product formation to investigate side-effects of NSAID treatment. This can be validated with extensive studies of NSAID use in the literature.

Goal 2:

a. Comparison of gene expression data with model variables for enzyme control to investigate how NSAIDs affect control of biological systems. Establish if perturbed regulation is related to scenarios (a)-(c) stated above.

b. Use data mining approach to determine how metabolic objective function changes in metabolic systems that are affected by drugs.

10.3 Final Thoughts

The recommended has great potential to transform medicinal development and application using cybernetic models and bioinformatics. While much data can be collected from biological systems, there is a dire lack of causality based theories that can predict what is going on at the systems level of an organism. Metabolite-drug interactions are no exception. The details of metabolic pathways are well characterized, but how these reactions are regulated as coordinated by the cell's genes is still unknown. Cybernetic models show great promise in providing this valuable insight and have the potential to ultimately elucidate the complex, synchronized behavior of intricate metabolic networks. This would constitute a significant triumph in the field

of bioinformatics and pave the way for the more purposeful development of drugs that interact with metabolism. When a drug is designed to affect the formation of some metabolite, a comprehensive understanding of how the metabolic system is affected will be known.

Each and every individual on the planet has the potential to benefit from the fruits borne from targeted drug design and applications. Understanding how drugs affect the metabolic process is crucial to understanding a drug's efficacy. Knowing the weaknesses in a cancer cell's metabolism through the formulation of cybernetic models can prevent the growth and evolution of the disease via the specification of appropriate drugs. Cybernetic models have already contributed much to the field of metabolic engineering. Their capacity to revolutionize treatment in pharmacological applications has yet to be realized. The recommended research has the capability of impacting a range of important activities in the application of medicine. Drug targets, treatment timing, dosage, and prediction of side effects could all be better informed by the cybernetic approach.

REFERENCES

REFERENCES

[1] Jacques Monod. The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3(1):371–394, 1949.

[2] P. Dhurjati, D. Ramkrishna, M. C. Flickinger, and G. T. Tsao. A cybernetic view of microbial growth: Modeling of cells as optimal strategists. *Biotechnology and Bioengineering*, 27(1):1–9, January 1985.

[3] Dhinakar S Kompala, Doraiswami Ramkrishna, Norman B Jansen, and George T Tsao. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. *Biotechnology and Bioengineering*, 28:1044–1055, 1986.

[4] Brian G. Turner and Doraiswami Ramkrishna. Revised enzyme synthesis rate expression in cybernetic models of bacterial growth. *Biotechnology and Bioengineering*, 31(1):41–43, January 1988.

[5] Brian G. Turner, D. Ramkrishna, and Norman B. Jansen. Cybernetic modeling of bacterial cultures at low growth rates: Mixed-substrate systems. *Biotechnology and Bioengineering*, 32(1):46–54, June 1988.

[6] S. Baloo and D. Ramkrishna. Metabolic regulation in bacterial continuous cultures: II. *Biotechnology and Bioengineering*, 38(11):1353–1363, December 1991.

[7] S Baloo and D Ramkrishna. Metabolic regulation in bacterial continuous cultures: I. *Biotechnology and bioengineering*, 38(11):1337–1352, 1991.

[8] Atul Narang, Allan Konopka, and D. Ramkrishna. New patterns of mixed-substrate utilization during batch growth of Escherichia coli K12. *Biotechnology and Bioengineering*, 55(5):747–757, 1997.

[9] Ramprasad Ramakrishna, Doraiswami Ramkrishna, and Allan E. Konopka. Cybernetic modeling of growth in mixed, substitutable substrate environments: Preferential and simultaneous utilization. *Biotechnology and Bioengineering*, 52(1):141–151, October 1996.

[10] Jeffrey V. Straight and Doraiswami Ramkrishna. Cybernetic Modeling and Regulation of Metabolic Pathways. Growth on Complementary Nutrients. *Biotechnology Progress*, 10(6):574–587, November 1994.

[11] J. Varner and D. Ramkrishna. Metabolic Engineering from a Cybernetic Perspective. 1. Theoretical Preliminaries. *Biotechnology Progress*, 15(3):407–425, January 1999.

[12] J. Varner and D. Ramkrishna. The non-linear analysis of cybernetic models. Guidelines for model formulation. *Journal of Biotechnology*, 71(13):67–103, May 1999.

[13] J. Varner and D. Ramkrishna. Metabolic Engineering from a Cybernetic Perspective: Aspartate Family of Amino Acids. *Metabolic Engineering*, 1(1):88–116, January 1999.

[14] Abhijit Namjoshi, Achim Kienle, and Doraiswami Ramkrishna. Steady-state multiplicity in bioreactors: bifurcation analysis of cybernetic models. *Chemical Engineering Science*, 58(36):793–800, February 2003.

[15] Abhijit Anand Namjoshi, Wei-Shou Hu, and Doraiswami Ramkrishna. Unveiling steady-state multiplicity in hybridoma cultures: The cybernetic approach. *Biotechnology and Bioengineering*, 81(1):80–91, January 2003.

[16] Jamey Young, Kristene Henne, John Morgan, Allan Konopka, and Doraiswami Ramkrishna. Cybernetic modeling of metabolism: towards a framework for rational design of recombinant organisms. *Chemical Engineering Science*, 59(2223):5041–5049, November 2004.

[17] Jamey D. Young, Kristene L. Henne, John A. Morgan, Allan E. Konopka, and Doraiswami Ramkrishna. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. *Biotechnology and Bioengineering*, 100(3):542–559, June 2008.

[18] Stefan Schuster, David A. Fell, and Thomas Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(3):326–332, March 2000.

[19] Jin Il Kim, Jeffery D. Varner, and Doraiswami Ramkrishna. A hybrid model of anaerobic E. coli GJT001: Combination of elementary flux modes and cybernetic variables. *Biotechnology Progress*, 24(5):993–1006, September 2008.

[20] Jin Il Kim. A hybrid cybernetic modeling for the growth of Escherichia coli in glucose-pyruvate mixtures. *Theses and Dissertations Available from ProQuest*, pages 1–185, January 2008.

[21] Jin Il Kim, Hyun-Seob Song, Sunil R. Sunkara, Arvind Lali, and Doraiswami Ramkrishna. Exacting predictions by cybernetic model confirmed experimentally: Steady state multiplicity in the chemostat. *Biotechnology Progress*, 28(5):1160–1166, September 2012.

[22] Hyun-Seob Song and Doraiswami Ramkrishna. Reduction of a set of elementary modes using yield analysis. *Biotechnology and Bioengineering*, 102(2):554–568, February 2009.

[23] Hyun-Seob Song and Doraiswami Ramkrishna. Prediction of metabolic function from limited data: Lumped hybrid cybernetic modeling (L-HCM). *Biotechnology and Bioengineering*, 106(2):271–284, 2010.

[24] Hyun-Seob Song and Doraiswami Ramkrishna. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnology and Bioengineering*, 108(1):127–140, 2011.

[25] Hyun-Seob Song and Doraiswami Ramkrishna. Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metabolic Engineering*, 14(2):69–80, March 2012.

[26] HS Song, D Ramkrishna, GE Pinchuk, AS Beliaev, AE Konopka, and JK Fredrickson. Dynamic Modeling of Aerobic Growth of Shewanella oneidensis. Predicting Triauxic Growth, Flux Distributions, and Energy Requirement for Growth. *Metabolic engineering*, 2012.

[27] Jacques Monod, Jean-Pierre Changeux, and Franois Jacob. Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 6(4):306–329, April 1963.

[28] Pablo A. Iglesias and Brian P. Ingalls. *Control Theory and Systems Biology*. MIT Press, 2010.

[29] Oliver Kotte, Judith B Zaugg, and Matthias Heinemann. Bacterial adaptation through distributed sensing of metabolic fluxes. *Molecular systems biology*, 6, 2010.

[30] Fiona A. Chandra, Gentian Buzi, and John C. Doyle. Glycolytic Oscillations and Limits on Robust Efficiency. *Science*, 333(6039):187–192, July 2011.

[31] Hannes Link, Dimitris Christodoulou, and Uwe Sauer. Advancing metabolic models with kinetic information. *Current Opinion in Biotechnology*, 29:8–14, October 2014.

[32] Linh M. Tran, Matthew L. Rizk, and James C. Liao. Ensemble Modeling of Metabolic Networks. *Biophysical Journal*, 95(12):5606–5617, December 2008.

[33] A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology*, 60(10):3724–3731, October 1994.

[34] Joanne M. Savinell and Bernhard O. Palsson. Optimal selection of metabolic fluxes for in vivo measurement. II. Application to Escherichia coli and hybridoma cell metabolism. *Journal of Theoretical Biology*, 155(2):215–242, March 1992.

[35] Daniel Segr, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117, November 2002.

[36] Tomer Shlomi, Omer Berkman, and Eytan Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7695–7700, May 2005.

[37] Robert Schuetz, Nicola Zamboni, Mattia Zampieri, Matthias Heinemann, and Uwe Sauer. Multidimensional Optimality of Microbial Metabolism. *Science*, 336(6081):601–604, May 2012.

[38] Aarash Bordbar, Jonathan M. Monk, Zachary A. King, and Bernhard O. Palsson. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, 15(2):107–120, February 2014.

[39] Radhakrishnan Mahadevan, Jeremy S. Edwards, and Francis J. Doyle III. Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, 83(3):1331–1340, September 2002.

[40] E Voit. Modelling metabolic networks using power-laws and S-systems, December 2008.

[41] Chih-Lung Ko, Feng-Sheng Wang, Yun-Peng Chao, and Te-Wei Chen. S-system approach to modeling recombinant Escherichia coli growth by hybrid differential evolution with data collocation. *Biochemical Engineering Journal*, 28(1):10–16, February 2006.

[42] H Kacser and JA Burns. The control of flux. *Symp. Soc. Exp. Biol.*, 27:65–104, 1973.

[43] George Stephanopoulos, Aristos A. Aristidou, and Jens Nielsen. *Metabolic Engineering: Principles and Methodologies*. Academic Press, October 1998.

[44] Peter D Grnwald. *The minimum description length principle*. MIT press, 2007.

[45] Hyun-Seob Song, Frank DeVilbiss, and Doraiswami Ramkrishna. Modeling metabolic systems: the need for dynamics. *Current Opinion in Chemical Engineering*, 2(4):373–382, November 2013.

[46] Jeffrey D. Orth, Ines Thiele, and Bernhard . Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, March 2010.

[47] M. Malik and A. Abdullah. A comparative study between flux balance analysis and kinetic model for C. acetobutylicum. In *Software Engineering Conference (MySEC), 2014 8th Malaysian*, pages 264–267, September 2014.

[48] Robert Schuetz, Lars Kuepfer, and Uwe Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, 3(1):119, January 2007.

[49] Hyun-Seob Song and Doraiswami Ramkrishna. When is the Quasi-Steady-State Approximation Admissible in Metabolic Modeling? When Admissible, What Models are Desirable? *Industrial & Engineering Chemistry Research*, 48(17):7976–7985, September 2009.

[50] C. E. Shannon. A Mathematical Theory of Communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001.

[51] Andrew Barron, Jorma Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *Information Theory, IEEE Transactions on*, 44:2743–2760, 1998.

[52] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[53] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6:461–464, 1978.

[54] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[55] Mark H Hansen and Bin Yu. Model Selection and the Principle of Minimum Description Length. *Journal of the American Statistical Association*, 96(454):746–774, June 2001.

[56] Xueyang Feng, You Xu, Yixin Chen, and Yinjie J. Tang. Integrating Flux Balance Analysis into Kinetic Models to Decipher the Dynamic Metabolism of Shewanella oneidensis MR-1. *PLoS Comput Biol*, 8:e1002376, 2012.

[57] Cong T Trinh, Pornkamol Unrean, and Friedrich Srienc. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. *Applied and environmental microbiology*, 74:3634–3643, 2008.

[58] Min-Kyu Oh, Lars Rohlin, Katy C. Kao, and James C. Liao. Global expression profiling of acetate-grown Escherichia coli. *The Journal of Biological Chemistry*, 277(15):13175–13183, April 2002.

[59] Yinjie J. Tang, Hector Garcia Martin, Adam Deutschbauer, Xueyang Feng, Rick Huang, Xavier Llora, Adam Arkin, and Jay D. Keasling. Invariability of central metabolic flux distribution in Shewanella oneidensis MR-1 under environmental or genetic perturbations. *Biotechnology Progress*, 25(5):1254–1259, 2009.

[60] L. Overbergh, A. Giulietti, D. Valckx, B. Decallonne, R. Bouillon, and C. Mathieu. The Use of Real-Time Reverse Transcriptase PCR for the Quantification of Cytokine Gene Expression. *Journal of Biomolecular Techniques : JBT*, 14(1):33–43, March 2003.

[61] Taesung Park, Sung-Gon Yi, Sung-Hyun Kang, SeungYeoun Lee, Yong-Sung Lee, and Richard Simon. Evaluation of normalization methods for microarray data. *BMC Bioinformatics*, 4(1):33, September 2003.

[62] Alan J. Wolfe. The Acetate Switch. *Microbiology and Molecular Biology Reviews*, 69(1):12–50, March 2005.

[63] Mansi El-Mansi, Alain J Cozzone, Joseph Shiloach, and Bernhard J Eikmanns. Control of carbon flux through enzymes of central and intermediary metabolism during growth of Escherichia coli on acetate. *Current Opinion in Microbiology*, 9(2):173–179, April 2006.

[64] Georges N. Cohen and Jacques Monod. BACTERIAL PERMEASES. *Bacteriological Reviews*, 21(3):169–194, September 1957.

[65] Katy C. Kao, Linh M. Tran, and James C. Liao. A Global Regulatory Role of Gluconeogenic Genes in Escherichia coli Revealed by Transcriptome Network Analysis. *Journal of Biological Chemistry*, 280(43):36079–36087, October 2005.

[66] Dong-Eun Chang, Darren J. Smalley, and Tyrrell Conway. Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model. *Molecular Microbiology*, 45(2):289–306, July 2002.

[67] Brice Enjalbert, Fabien Letisse, and Jean-Charles Portais. Physiological and Molecular Timing of the Glucose to Acetate Transition in Escherichia coli. *Metabolites*, 3(3):820–837, September 2013.

[68] G. N. Vemuri, E. Altman, D. P. Sangurdekar, A. B. Khodursky, and M. A. Eiteman. Overflow Metabolism in Escherichia coli during Steady-State Growth: Transcriptional Regulation and Effect of the Redox Ratio. *Applied and Environmental Microbiology*, 72(5):3653–3661, May 2006.

[69] Chotirat Ann Ralanamahatana, Jessica Lin, Dimitrios Gunopulos, Eamonn Keogh, Michail Vlachos, and Gautam Das. Mining Time Series Data. In Oded Maimon and Lior Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pages 1069–1103. Springer US, January 2005.

[70] Jiao Zhao and Kazuyuki Shimizu. Metabolic flux analysis of Escherichia coli K12 grown on 13c-labeled acetate and glucose using GC-MS and powerful flux calculation method. *Journal of Biotechnology*, 101(2):101–117, March 2003.

[71] Annik Perrenoud and Uwe Sauer. Impact of Global Transcriptional Regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on Glucose Catabolism in Escherichia coli. *Journal of Bacteriology*, 187(9):3171–3179, May 2005.

[72] Mehdi Rahimpour, Manuel Montero, Goizeder Almagro, Alejandro M. Viale, ngel Sevilla, Manuel Cnovas, Francisco J. Muoz, Edurne Baroja-Fernndez, Abdellatif Bahaji, Gustavo Eydallin, Hitomi Dose, Rikiya Takeuchi, Hirotada Mori, and Javier Pozueta-Romero. GlgS, described previously as a glycogen synthesis control protein, negatively regulates motility and biofilm formation in Escherichia coli. *The Biochemical Journal*, 452(3):559–573, June 2013.

[73] M. Freundlich, N. Ramani, E. Mathew, A. Sirko, and P. Tsui. The role of integration host factor In gene expression in Escherichia coli. *Molecular Microbiology*, 6(18):2557–2563, September 1992.

[74] Y. Maki, H. Yoshida, and A. Wada. Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase Escherichia coli. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 5(12):965–974, December 2000.

[75] J. B. Stock, E. B. Waygood, N. D. Meadow, P. W. Postma, and S. Roseman. Sugar transport by the bacterial phosphotransferase system. The glucose receptors of the Salmonella typhimurium phosphotransferase system. *Journal of Biological Chemistry*, 257(23):14543–14552, December 1982.

[76] Kirill Peskov, Ekaterina Mogilevskaya, and Oleg Demin. Kinetic modelling of central carbon metabolism in Escherichia coli. *FEBS Journal*, 279:3374–3385, 2012.

[77] Jacqueline Plumbridge. Expression of ptsG, the gene for the major glucose PTS transporter in Escherichia coli, is repressed by Mlc and induced by growth on glucose. *Molecular Microbiology*, 29(4):1053–1063, August 1998.

[78] Shona Seeto, Lucinda Notley-McRobb, and Thomas Ferenci. The multifactorial influences of RpoS, Mlc and cAMP on ptsG expression under glucose-limited and anaerobic conditions. *Research in Microbiology*, 155(3):211–215, April 2004.

[79] Dongwoo Shin, Namwook Cho, Sunggi Heu, and Sangryeol Ryu. Selective regulation of ptsG expression by Fis. Formation of either activating or repressing nucleoprotein complex in response to glucose. *The Journal of Biological Chemistry*, 278(17):14776–14781, April 2003.

[80] Aravinda R. Mandli and Jayant M. Modak. Cybernetic modeling of adaptive prediction of environmental changes by microorganisms. *Mathematical Biosciences*, 248:40–45, February 2014.

[81] Emanuela Ricciotti and Garret A. FitzGerald. Prostaglandins and Inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(5):986–1000, May 2011.

[82] Gary A. Green. Understanding NSAIDs: From aspirin to COX-2. *Clinical Cornerstone*, 3(5):50–59, 2001.

[83] Shakti Gupta, Mano Ram Maurya, Daren L Stephens, Edward A Dennis, and Shankar Subramaniam. An integrated model of eicosanoid metabolism and signaling based on lipidomics flux analysis. *Biophysical journal*, 96:4542–4551, 2009.

[84] Yasuyuki Kihara, Shakti Gupta, Mano R. Maurya, Aaron Armando, Ishita Shah, Oswald Quehenberger, Christopher K. Glass, Edward A. Dennis, and Shankar Subramaniam. Modeling of Eicosanoid Fluxes Reveals Functional Coupling between Cyclooxygenases and Terminal Synthases. *Biophysical Journal*, 106(4):966–975, February 2014.

[85] Kun Yang, Wenzhe Ma, Huanhuan Liang, Qi Ouyang, Chao Tang, and Luhua Lai. Dynamic Simulations on the Arachidonic Acid Metabolic Network. *PLoS Comput Biol*, 3(3):e55, March 2007.

[86] Chong He, Yiran Wu, Yongquan Lai, Zongwei Cai, Ying Liu, and Luhua Lai. Dynamic eicosanoid responses upon different inhibitor and combination treatments on the arachidonic acid metabolic network. *Molecular BioSystems*, 8(5):1585–1594, April 2012.

[87] Christian R. H. Raetz, Teresa A. Garrett, C. Michael Reynolds, Walter A. Shaw, Jeff D. Moore, Dale C. Smith, Anthony A. Ribeiro, Robert C. Murphy, Richard J. Ulevitch, Colleen Fearns, Donna Reichart, Christopher K. Glass, Chris Benner, Shankar Subramaniam, Richard Harkewicz, Rebecca C. Bowers-Gentry, Matthew W. Buczynski, Jennifer A. Cooper, Raymond A. Deems, and Edward A. Dennis. Kdo2-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4. *Journal of Lipid Research*, 47(5):1097–1111, May 2006.

[88] Mausumee Guha and Nigel Mackman. LPS induction of gene expression in human monocytes. *Cellular Signalling*, 13(2):85–94, February 2001.

[89] Edward A. Dennis, Raymond A. Deems, Richard Harkewicz, Oswald Quehenberger, H. Alex Brown, Stephen B. Milne, David S. Myers, Christopher K. Glass, Gary Hardiman, Donna Reichart, Alfred H. Merrill, M. Cameron Sullards, Elaine Wang, Robert C. Murphy, Christian R. H. Raetz, Teresa A. Garrett, Ziqiang Guan, Andrea C. Ryan, David W. Russell, Jeffrey G. McDonald, Bonne M. Thompson, Walter A. Shaw, Manish Sud, Yihua Zhao, Shakti Gupta, Mano R. Maurya, Eoin Fahy, and Shankar Subramaniam. A Mouse Macrophage Lipidome. *Journal of Biological Chemistry*, 285(51):39976–39985, December 2010.

[90] Hong Huang, Tongzheng Liu, Jane L. Rose, Rachel L. Stevens, and Dale G. Hoyt. Sensitivity of mice to lipopolysaccharide is increased by a high saturated fat and cholesterol diet. *Journal of Inflammation*, 4(1):22, November 2007.

[91] Kazuhito Tsuboi, Yukihiko Sugimoto, and Atsushi Ichikawa. Prostanoid receptor subtypes. *Prostaglandins & other lipid mediators*, 68:535–556, 2002.

[92] Tjomme van der Bruggen, Suzanne Nijenhuis, Estia van Raaij, Jan Verhoef, and B. Sweder van Asbeck. Lipopolysaccharide-Induced Tumor Necrosis Factor Alpha Production by Human Monocytes Involves the Raf-1/MEK1-MEK2/ERK1-ERK2 Pathway. *Infection and Immunity*, 67(8):3824–3829, August 1999.

[93] Joachim Rothe, Werner Lesslauer, Hansruedi Ltscher, Yolande Lang, Pascale Koebel, Frank Kntgen, Alana Althage, Rolf Zinkernagel, Michael Steinmetz, and Horst Bluethmann. Mice lacking the tumour necrosis factor receptor 1 are resistant to IMF-mediated toxicity but highly susceptible to infection by Listeria monocytogenes. *Nature*, 364(6440):798–802, August 1993.

[94] N. S. Trede, A. V. Tsytsykova, T. Chatila, A. E. Goldfeld, and R. S. Geha. Transcriptional activation of the human TNF-alpha promoter by superantigen in human monocytic cells: role of NF-kappa B. *The Journal of Immunology*, 155(2):902–908, July 1995.

[95] Chandan K. Sen, Sashwati Roy, and Lester Packer. Involvement of intracellular Ca2+ in oxidant-induced NF-B activation. *FEBS Letters*, 390(2):241–241, July 1996.

[96] Takako Hirata and Shuh Narumiya. Prostanoid Receptors. *Chemical Reviews*, 111(10):6209–6230, October 2011.

[97] Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Kegg as a reference resource for gene and protein annotation. *Nucleic acids research*, 44(D1):D457–D462, 2016.

[98] Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 27(1):29–34, 1999.

[99] C. J. Grossman, J. Wiseman, F. S. Lucas, M. A. Trevethick, and P. J. Birch. Inhibition of constitutive and inducible cyclooxygenase activity in human platelets and mononuclear cells by NSAIDS and Cox 2 inhibitors. *Inflammation Research*, 44(6):253–257, June 1995.

[100] E. A. Meade, W. L. Smith, and D. L. DeWitt. Differential inhibition of prostaglandin endoperoxide synthase (cyclooxygenase) isozymes by aspirin and other non-steroidal anti-inflammatory drugs. *Journal of Biological Chemistry*, 268(9):6610–6614, March 1993.

[101] Frances R. Balkwill. *The Cytokine Network*. Oxford University Press, 2000.

[102] Philipp Adler, Hyun-Seob Song, Katharina Kstner, Doraiswami Ramkrishna, and Benno Kunz. Prediction of dynamic metabolic behavior of Pediococcus pentosaceus producing lactic acid from lignocellulosic sugars. *Biotechnology Progress*, 28(3):623–635, May 2012.

[103] Mehmet A. Orman, Francois Berthiaume, Ioannis P. Androulakis, and Marianthi G. Ierapetritou. Advanced Stoichiometric Analysis of Metabolic Networks of Mammalian Systems. *Critical Reviews in Biomedical Engineering*, 39(6):511–534, 2011.

[104] Eleftherios Pilalis, Aristotelis Chatziioannou, Brigitte Thomasset, and Fragiskos Kolisis. An in silico compartmentalized metabolic model of Brassica napus enables the systemic study of regulatory aspects of plant central metabolism. *Biotechnology and Bioengineering*, 108(7):1673–1682, July 2011.

[105] Neema Jamshidi and Bernhard . Palsson. Systems biology of the human red blood cell. *Blood Cells, Molecules, and Diseases*, 36(2):239–247, March 2006.

[106] Jamey D Young. Learning from the steersman: A natural history of cybernetic models. *Industrial & Engineering Chemistry Research*, 54(42):10162–10169, 2015.

[107] Anthony P. Burgard and Costas D. Maranas. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and Bioengineering*, 82(6):670–677, June 2003.

[108] Erwin P Gianchandani, Matthew A Oberhardt, Anthony P Burgard, Costas D Maranas, and Jason A Papin. Predicting biological system objectives de novo from internal state measurements. *BMC Bioinformatics*, 9:43, January 2008.

[109] Gautam Goel, I.-Chun Chou, and Eberhard O. Voit. System estimation from metabolic time-series data. *Bioinformatics*, 24(21):2505–2511, November 2008.

[110] Gengjie Jia, Gregory N. Stephanopoulos, and Rudiyanto Gunawan. Parameter estimation of kinetic models from metabolic profiles: two-phase dynamic decoupling method. *Bioinformatics*, 27(14):1964–1970, July 2011.

[111] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1):44–57, December 2008.

[112] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, October 2005.

[113] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1):1–13, January 2009.

[114] John L. Ingraham, Ole Maale, Frederick C. (Frederick Carl) Neidhardt, and 1931. *Growth of the bacterial cell*. Sinauer Associates, 1983.

[115] Jennifer L. Stow, Anthony P. Manderson, and Rachael Z. Murray. SNAREing immunity: the role of SNAREs in the immune system. *Nature Reviews Immunology*, 6(12):919–929, December 2006.

[116] Pei-Chin Chuang, Yiu-Juian Lin, Meng-Hsing Wu, Lih-Yuh C. Wing, Yutaka Shoji, and Shaw-Jenq Tsai. Inhibition of CD36-Dependent Phagocytosis by Prostaglandin E2 Contributes to the Development of Endometriosis. *The American Journal of Pathology*, 176(2):850–860, February 2010.

[117] Joyce P. Cox and Manfred L. Karnovsky. The Depression of Phagocytosis by Exogenous Cyclic Nucleotides, Prostaglandins, and Theophylline. *The Journal of Cell Biology*, 59(2):480–490, November 1973.

[118] Dr Robert D. Prusch, Stella-Maris Goette, and Paula Haberman. Prostaglandins may play a signal-coupling role during phagocytosis in Amoeba proteus. *Cell and Tissue Research*, 255(3):553–557, March 1989.

[119] Lisa M. Rogers, Tennille Thelen, Krystle Fordyce, Emilie Bourdonnay, Casey Lewis, Han Yu, Junyong Zhang, Jingli Xie, Carlos H. Serezani, Marc Peters-Golden, and David M. Aronoff. EP4 and EP2 Receptor Activation of Protein Kinase A by Prostaglandin E2 Impairs Macrophage Phagocytosis of Clostridium sordellii. *American Journal of Reproductive Immunology*, 71(1):34–43, January 2014.

[120] Markus Bruckner, Denise Dickel, Eva Singer, and Daniel F. Legler. Distinct modulation of chemokine expression patterns in human monocyte-derived dendritic cells by prostaglandin E2. *Cellular Immunology*, 276(12):52–58, March 2012.

[121] Eri Kawashita, Daisuke Tsuji, Masahiro Toyoshima, Yosuke Kanno, Hiroyuki Matsuno, and Kohji Itoh. Prostaglandin E2 Reverses Aberrant Production of an Inflammatory Chemokine by Microglia from Sandhoff Disease Model Mice through the cAMP-PKA Pathway. *PLOS ONE*, 6(1):e16269, January 2011.

[122] Marie T. Rubio, Terry K. Means, Ronjon Chakraverty, Juanita Shaffer, Yasuhiro Fudaba, Meredith Chittenden, Andrew D. Luster, and Megan Sykes. Maturation of human monocyte-derived dendritic cells (MoDCs) in the presence of prostaglandin E2 optimizes CD4 and CD8 T cell-mediated responses to protein antigens: role of PGE2 in chemokine and cytokine expression by MoDCs. *International Immunology*, 17(12):1561–1572, December 2005.

[123] Kiyoshi Takayama, Guillermo Garca-Cardea, Galina K. Sukhova, Jason Comander, Michael A. Gimbrone, and Peter Libby. Prostaglandin E2 Suppresses Chemokine Production in Human Macrophages through the EP4 Receptor. *Journal of Biological Chemistry*, 277(46):44147–44154, November 2002.

[124] Paul C. Norris and Edward A. Dennis. A lipidomic perspective on inflammatory macrophage eicosanoid signaling. *Advances in Biological Regulation*, 54:99–110, January 2014.

[125] Hyun-Seob Song, John A. Morgan, and Doraiswami Ramkrishna. Systematic development of hybrid cybernetic models: Application to recombinant yeast co-consuming glucose and xylose. *Biotechnology and Bioengineering*, 103(5):984–1002, August 2009.

[126] Joachim Almquist, Marija Cvijovic, Vassily Hatzimanikatis, Jens Nielsen, and Mats Jirstrand. Kinetic models in industrial biotechnology Improving cell factory performance. *Metabolic Engineering*, 24:38–60, July 2014.

[127] Ida Schomburg, Antje Chang, Christian Ebeling, Marion Gremse, Christian Heldt, Gregor Huhn, and Dietmar Schomburg. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Research*, 32(suppl 1):D431–D433, January 2004.

[128] Yun Lee, Jimmy G. Lafontaine Rivera, and James C. Liao. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metabolic Engineering*, 25:63–71, September 2014.

[129] Timo Lubitz, Marvin Schulz, Edda Klipp, and Wolfram Liebermeister. Parameter Balancing in Kinetic Models of Cell Metabolism. *The Journal of Physical Chemistry B*, 114(49):16298–16303, December 2010.

[130] Sebastian Aljoscha Wahl, Katharina Nh, and Wolfgang Wiechert. 13c labeling experiments at metabolic nonstationary conditions: An exploratory study. *BMC Bioinformatics*, 9:152, 2008.

[131] Carmen G. Moles, Pedro Mendes, and Julio R. Banga. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Research*, 13(11):2467–2474, November 2003.

[132] M. Srinivas and L. M. Patnaik. Genetic algorithms: a survey. *Computer*, 27(6):17–26, June 1994.

[133] Robert Hooke and T. A. Jeeves. Direct Search Solution of Numerical and Statistical Problems. *J. ACM*, 8(2):212–229, April 1961.

[134] Novartis Foundation. *'In Silico' Simulation of Biological Processes*. John Wiley & Sons, July 2003.

[135] A. Mardinoglu and J. Nielsen. Systems medicine and metabolic modelling. *Journal of Internal Medicine*, 271(2):142–154, February 2012.

[136] Jun Zou, Ming-Wu Zheng, Gen Li, Zhi-Guang Su, Jun Zou, Ming-Wu Zheng, Gen Li, and Zhi-Guang Su. Advanced Systems Biology Methods in Drug Discovery and Translational Biomedicine, Advanced Systems Biology Methods in Drug Discovery and Translational Biomedicine. *BioMed Research International, BioMed Research International*, 2013, 2013:e742835, September 2013.

[137] J. Bruce German, Matthew-Alan Roberts, and Steven M. Watkins. Genomics and Metabolomics as Markers for the Interaction of Diet and Health: Lessons from Lipids. *The Journal of Nutrition*, 133(6):2078S–2083S, June 2003.

[138] O. Folger, L. Jerby, C. Frezza, E. Gottlieb, E. Ruppin, and T. Shlomi. Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology*, 7(1):501–501, April 2014.

[139] Christian Frezza, Liang Zheng, Ori Folger, Kartik N. Rajagopalan, Elaine D. MacKenzie, Livnat Jerby, Massimo Micaroni, Barbara Chaneton, Julie Adam, Ann Hedley, Gabriela Kalna, Ian P. M. Tomlinson, Patrick J. Pollard, Dave G. Watson, Ralph J. Deberardinis, Tomer Shlomi, Eytan Ruppin, and Eyal Gottlieb. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*, 477(7363):225–228, September 2011.

[140] Jonathan R. Karr, Jayodita C. Sanghvi, Derek N. Macklin, Miriam V. Gutschow, Jared M. Jacobs, Benjamin Bolival, Nacyra Assad-Garcia, John I. Glass, and Markus W. Covert. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, 150(2):389–401, July 2012.

APPENDICES

A. DEVELOPMENT OF METABOLIC MODELS FOR MODEL SELECTION TEST CASES

This segment features relevant details related to the development of the dynamic models used in the dynamic model comparison in the preliminary results section.

A.1 Network

The reaction network used in this model selection exercise was taken from Kotte et al. (2011) from which the figure is taken and placed below. On the right side of the figure, the metabolic reactions involved are shown. In the figure within the main body of the report, the reactions correspond to the reaction names below. The other information related to the kinetic model is featured in table A.1 including the influences of different signaling mechanisms.

Table A.1.: Reaction Network

Reaction Name	Balance	Reversible
Biomass	3.75 AcCoA + 5.20 alKG + 1.33 G6P + 18.27 NADPH + 1.79 OxA + 0.52 PEP + 1.50 PG3 + 2.83 Pyr + 59.89 ATP = 1 Biomass + 3.55 NADH	
Uptake/Exchange		
Glc_PTS_up	1 Gluc_ext + 1 PEP = 1 G6P + 1 Pyr	
Ac_up	1 Ac_ext = 1 Ac	
Ac_out	1 Ac = 1 Ac_ext	

continued on next page

Table A.1.: *continued*

Reaction Name	Balance	Reversible
O2_up	1 O2_ext = 1 O2	
CO2_ex	1 CO2_ext = 1 CO2	Yes
Glycolysis		
G6P::FBP	1 G6P + 1 ATP = 1 FBP	
FBP::G6P	1 FBP = 1 G6P	
FBP::PG3	1 FBP = 2 PG3 + 2 NADH + 2 ATP	Yes
PG3::PEP	1 PG3 = 1 PEP	Yes
PEP::Pyr	1 PEP = 1 Pyr + 1 ATP	
Pyr::PEP	1 Pyr + 2 ATP = 1 PEP	
Pyr::AcCoA	1 Pyr = 1 AcCoA + 1 NADH + 1 CO	
TCA		
AcCoA::ICit	1 AcCoA + 1 OxA = 1 ICit	
ICit::alKG	1 ICit = 1 alKG + NADH + CO2	
alKG::Mal	1 alKG = 1 Mal + 1 ATP + 1 NADH + 1 CO2	
Mal::OxA	1 Mal = 1 OxA + NADH	
OxA::PEP	1 OxA + 1 ATP = 1 PEP + 1 CO2	
PEP::OxA	1 PEP + 1 CO2 = 1 OxA	
Acetate Metabolism		
AcCoA::AcP_r	1 AcCoA = 1 Ac	Yes
Mal::Pyr	1 Mal = 1 Pyr + 1 NADH + 1 CO2	
Ac::AcCoA	2 ATP + 1 Ac = 1 AcCoA	
Glyox		
ICit::Glyox	1 ICit = 1 alKG + 1 Glyox	
Glyox::Mal	1 AcCoA + 1 Glyox = 1 Mal	
Oxid Phos		

continued on next page

Table A.1.: *continued*

Reaction Name	Balance	Reversible
Oxidase	1 QH2 + 0.5 O2 = 2 Hp	
NADHDehydro	1 NADH = 1 QH2 + 2 Hp	Yes
TransHydro	1 NADH + 1 Hp = 1 NADPH	Yes
ATPSynth	3 Hp = 1 ATP	Yes
Maintenance		
ATPdrain	1 ATP = 1 ATP_ext	

A slightly more detailed network was used to develop the dFBA, HCM and L-HCM models. These networks include both structured biomass equations and maintenance reactions. The structured biomass equation is a simplification of the one found in (Trinh et al. 2008).

A.2 Development of dFBA Description

The dFBA model was formed using a Monod type uptake kinetics which have the following formulation:

$$\frac{dX}{dt} = X(\mu_G + \mu_A)$$
$$\frac{dG}{dt} = -X\mu_G/Y_G - r_{A,G}$$
$$\frac{dA}{dt} = r_{A,G} - X\mu_A/Y_A$$

where

$$\mu_G = \frac{\mu_G^{\max} G}{K_G + G}$$
$$\mu_A = \frac{\mu_A^{\max} A}{K_A + A}.$$

Table A.2.: Parameters Used for dFBA Model

Parameter	Value
μ_G^{max}	0.60898005393401899
K_G	5.0099999999999999E-2
μ_G^{max}	1.50317424025733E-2
K_A	0.79990000000000006
k_A	0.41966587761481799
Y_G	0.38668959071425302
Y_A	4.4608009913101997E-3

In this case, X, G, and A denote the concentrations of biomass, glucose and acetate respectively. Glucose is taken up by the biomass and converted to acetate which is represented by the term $r_{A,G}$.

$$r_{A,G} = k_A GX$$

The model outlined has a total of 7 different parameters which take on the values listed in table A.2.

The parameters above were solved for using Matlabs GA and fminsearch functions in order to minimize a sum of squared errors function. Figure A.1 shows the dynamic behavior of the 3 differential equations and their closeness to the behavior modeled.

From the Monod model, the rates of substrate uptake were inferred to use in FBA to arrive at dynamic flux behaviors. The FBA problem used biomass maximization as an objective function and was calculated by solving a linear programming problem.

By dividing the dynamic uptake data into discrete intervals, one can infer the dynamic flux solution through multiple linear program solutions using the different uptake rates.

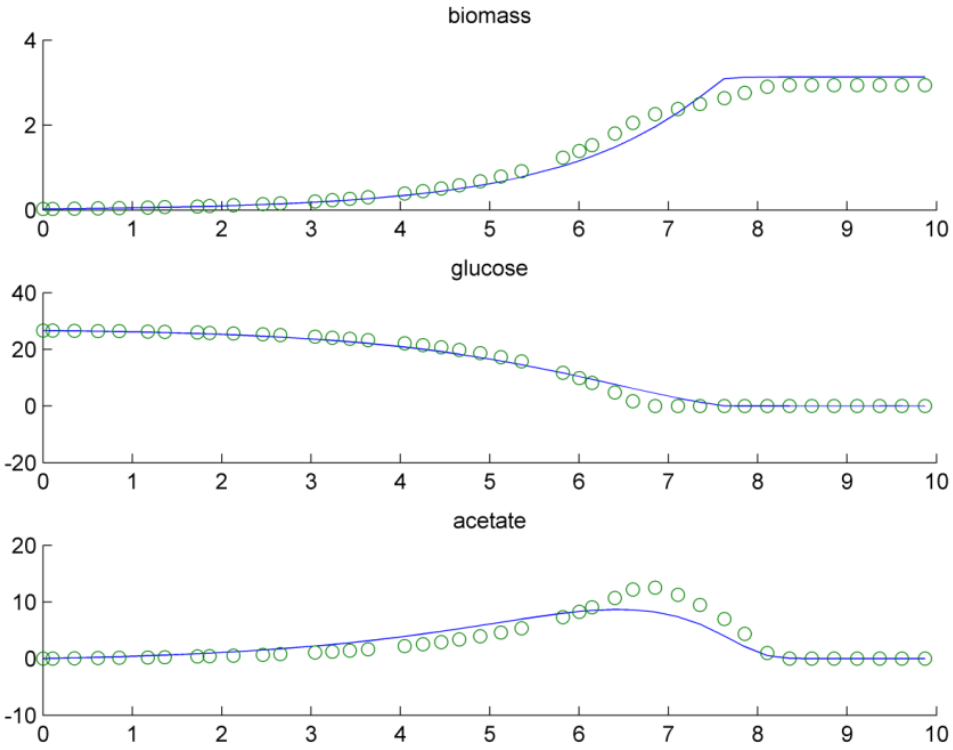

Figure A.1.: Visualization of Monod model for dFBA uptake rates.

A.3 HCM and L-HCM Model Development

The same reaction network as above was used for both the L-HCM and HCM models. For the development of both models, EFMs were calculated for the reaction network. The total number of EFMs amounted to120 total. For the HCM model, it was infeasible to specify 240 parameters (k_i and (K_i for each mode). Instead, yield analysis (YA) (Song et al. 2009) was used to reduce the total number of elementary modes by determining which elementary modes made up the extreme edges of the yield space. Yield space was calculated around the glucose and acetate uptake/excretion. Elementary modes at the extreme edges were selected to preserve 99% of the n-dimensional volume. This process reduced the elementary modes from 120 to 7 modes. The modes are all yields are $Y_{glu/bio}$, $Y_{ac_{up/bi}o}$, and $Y_{ac_{out/bio}}$. Note that the inverse is more intuitive but undefined when the production of biomass in the absence of one of the uptake or production of a product. The total space analyzed is visualized in figure A.2.

These modes are listed in table A.3.

Table A.3.: Elementary Modes Isolated from Yield Analysis

Reaction Name	1	2	3	4	5	6	7
'G6P::FBP'	1.93	0	6.8	0	6.86	0.13	35.06
'PEP::Pyr'	0.45	0	3.93	0	0	0	24.4
'Pyr::PEP'	0	2	0	0	1.07	2	0
'Pyr::AcCoA'	2.16	0	9.89	6.26	7.16	0	54.67
'AcCoA::ICit'	0.74	1.92	2.91	9.88	5.08	2.02	22.26
'ICit::alKG'	0.74	1.75	0	0	5.08	2.02	0
'alKG::Mal'	0	1.86	0	8.58	2.21	1.9	5.69
'Mal::OxA'	0	0	2.91	11.49	0	0	27.95

continued on next page

Table A.3.: *continued*

Reaction Name	1	2	3	4	5	6	7
'OxA::PEP'	0	0	0	1.17	0	0	0
'PEP::OxA'	1	1.94	1	0	6.07	2.06	0
'ICit::Glyox'	0	0.17	2.91	9.88	0	0	22.26
'Glyox::Mal'	0	0.17	2.91	9.88	0	0	22.26
'Oxidase'	4.65	5.43	18.16	28.96	22.21	5.75	111.51
'AcCoA::Ac'	0.88	0	1.98	0	0	0	0
'Ac::AcCoA'	0	2.14	0	14.44	0	2.1	1.78
'ATPdrain'	0	0	0	0	0	0	0
'mue'	0.14	0.01	0.56	0.25	0.55	0.02	3.19
'Glc_PTS_up'	2.12	0	7.55	0	7.59	0.16	39.29
'O2_up'	2.32	2.72	9.08	14.48	11.11	2.88	55.76
'Mal::Pyr'	0	2.03	0	6.97	2.21	1.9	0
'FBP::G6P'	0	0.02	0	0.33	0	0	0
'FBP::PG3'	1.93	-0.02	6.8	-0.33	6.86	0.13	35.06
'PG3::PEP'	3.64	-0.05	12.77	-1.04	12.88	0.23	65.35
'NADHDehydro'	4.65	5.43	18.16	28.96	22.21	5.75	111.51
'TransHydro'	2.62	0.22	10.22	4.56	10.12	0.41	58.22
'ATPSynth'	5.32	7.17	20.81	37.09	26.25	7.53	129.27

For the generation of the HCM model, 10 parameters were fit to the data using MATLABs fmincon nonlinear optimization gradient search from 8 different starting points for robust parameter calculation. Seven k_i^{max} parameters were calculated for each mode. 2 Michaelis-Menten constants were calculated for both glucose and acetate substrates. 1 enzyme synthesis parameter was determined. All model constants are stated in table A.4.

The fit of the HCM model to external substrate data is shown in figure A.3.

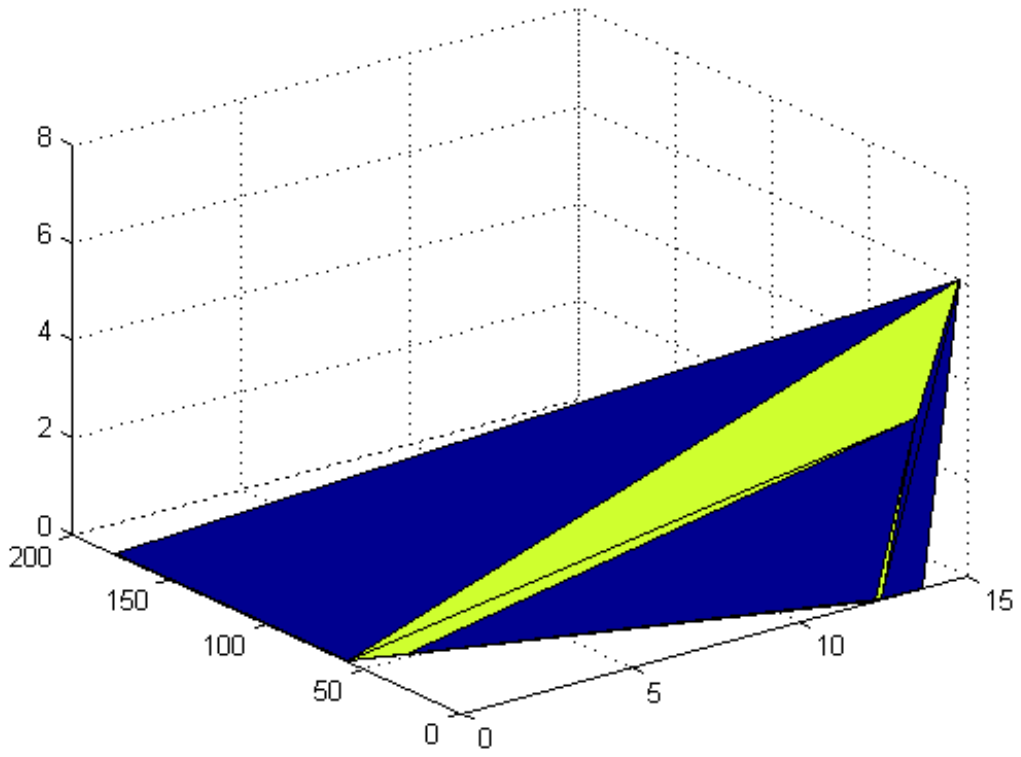

Figure A.2.: Convex hull used for yield analysis.

Table A.4.: HCM Model Parameters

Name	Parameter Values
k_1^{max}	32.27
k_2^{max}	1.2
k_3^{max}	4.81
k_4^{max}	1.67
k_5^{max}	5.16
k_6^{max}	26.61
k_7^{max}	1.66
K_{Glu}	1.77
K_{Ac}	0.43
k_E	1.49

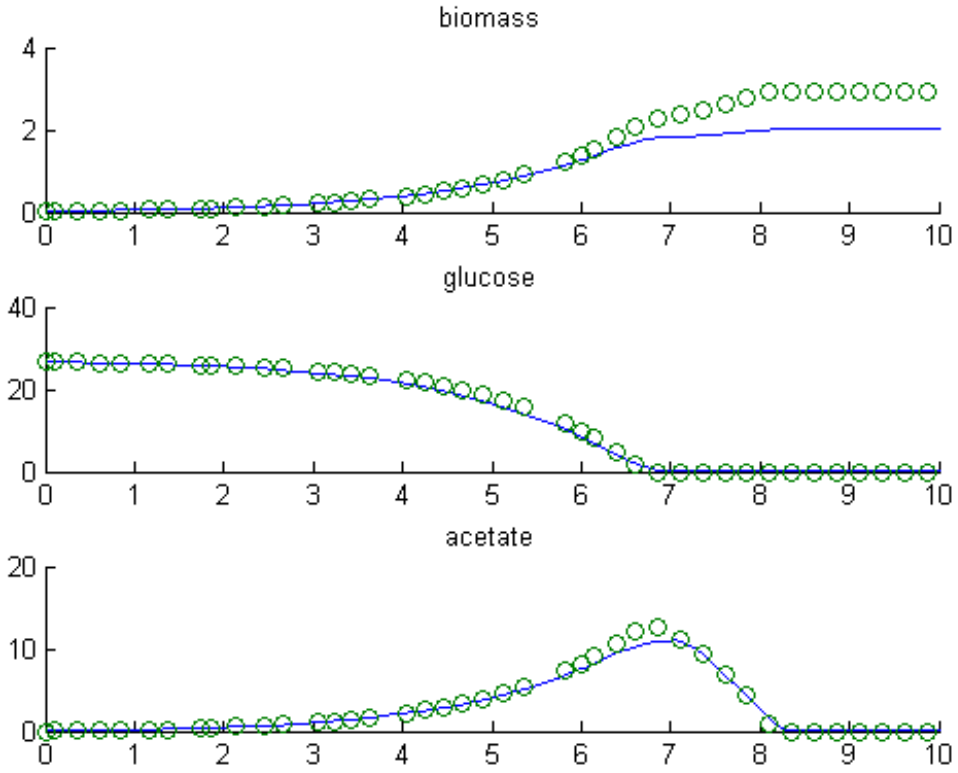

Figure A.3.: Visualization of HCM description of extracellular species.

The L-HCM model was generated using the same network. The metabolite data for the yields of biomass and acetate growing on glucose and the yield of biomass on acetate were determined by fitting logistic curves to the batch data. These yields were used to lump the modes into two families for growth on glucose and for growth on acetate. The yield data was used to tune the lumping parameter. This was done using a genetic algorithm and fminsearch to minimize the difference between the lumped modes yields and actual data for these yields. The tuning parameters n_η were 11.92 and 5.57 for the glucose and acetate modes respectively. The GAR used to weight maintenance energy was 97. The lumped modes are described in table A.5.

From these 2 lumped elementary modes, only two k_i^{max} parameters were fit to the data. Other parameters were assumed to be robust and were taken from other L-HCM work (Song et al. 2010/2011). The values for these parameters are in table A.6.

Table A.5.: L-HCM Family Modes

Reaction Name	Glucose	Acetate
'G6P::FBP'	0.95	0
'PEP::Pyr'	0.58	0.21
'Pyr::PEP'	3.50E-2	1.9E-2
'Pyr::AcCoA'	1.52	0.489
'AcCoA::ICit'	0.42	0.73
'ICit::alKG'	2.50E-2	2.3E-2
'alKG::Mal'	0.11	0.65
'OxA::PEP'	0	0.26
'PEP::OxA'	0.18	0
'ICit::Glyox'	0.39	0.70
'Glyox::Mal'	0.39	0.70
'Oxidase'	3.11	2.24
'AcCoA::Ac'	0.50	0
'Ac::AcCoA'	0	1
'ATPdrain'	2.214	0.52
'mue'	0.06	1.50E-2
'Glc_PTS_up'	1	0
'O2_up'	1.56	1.12
'Mal::Pyr'	0.151	0.34
'FBP::G6P'	0.03	0.02
'FBP::PG3'	0.92	-0.02
'PG3::PEP'	1.75	-6.2E-2
'Mal::OxA'	0.34	1.012
'NADHDehydro'	3.11	2.24
'TransHydro'	1.10	0.27
'ATPSynth'	3.79	2.9

Table A.6.: L-HCM Family Parameters

Name	Value
k_{glu}^{max}	10.73
k_{ac}^{max}	19.28

B. GENE EXPRESSION PROFILES FOR COMPLETE SET OF GENES FOR GLUCOSE-ACETATE SYSTEM

Description

The next few figures show the behavior of different gene transcripts with their respective cybernetic variables for different pathways as the cells shift from glucose to acetate consumption. u_i variables for comparison are selected as indicated in body of the dissertation in chapter 4.

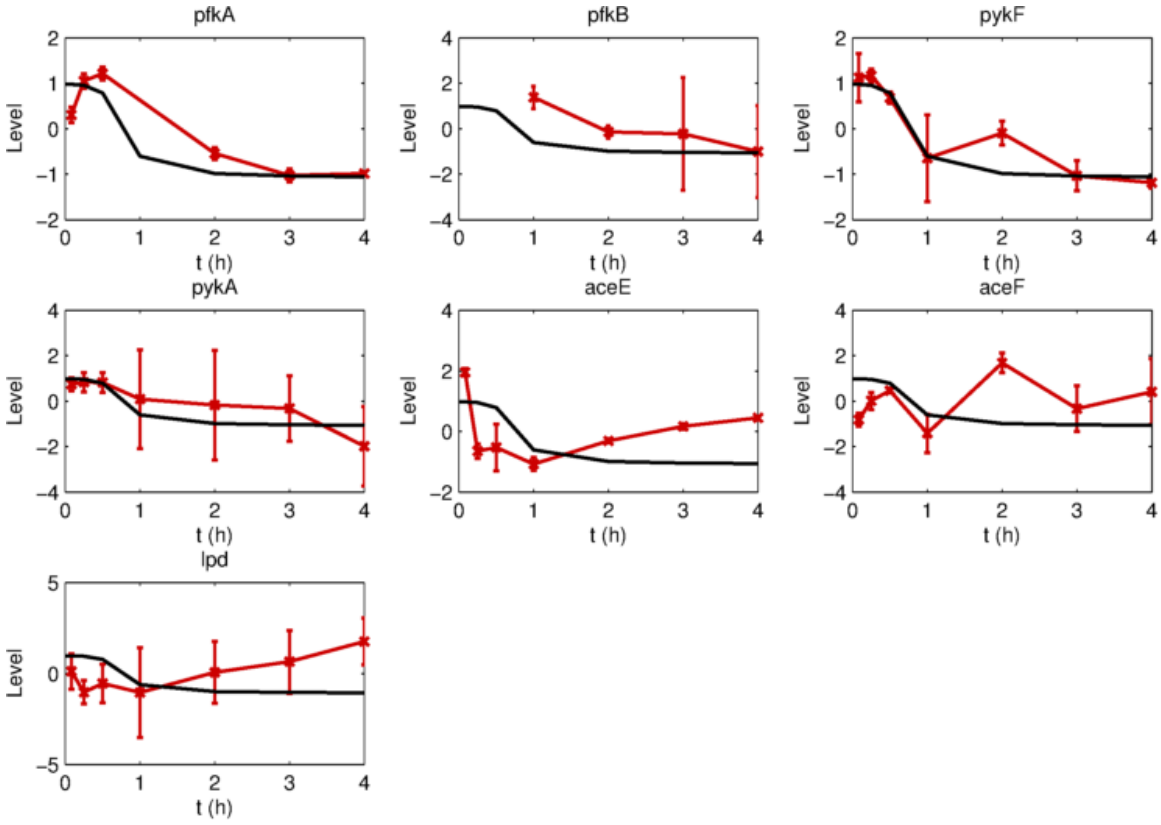

Figure B.1.: Glycolysis pathway gene transcripts compared to glucose u_i variable.

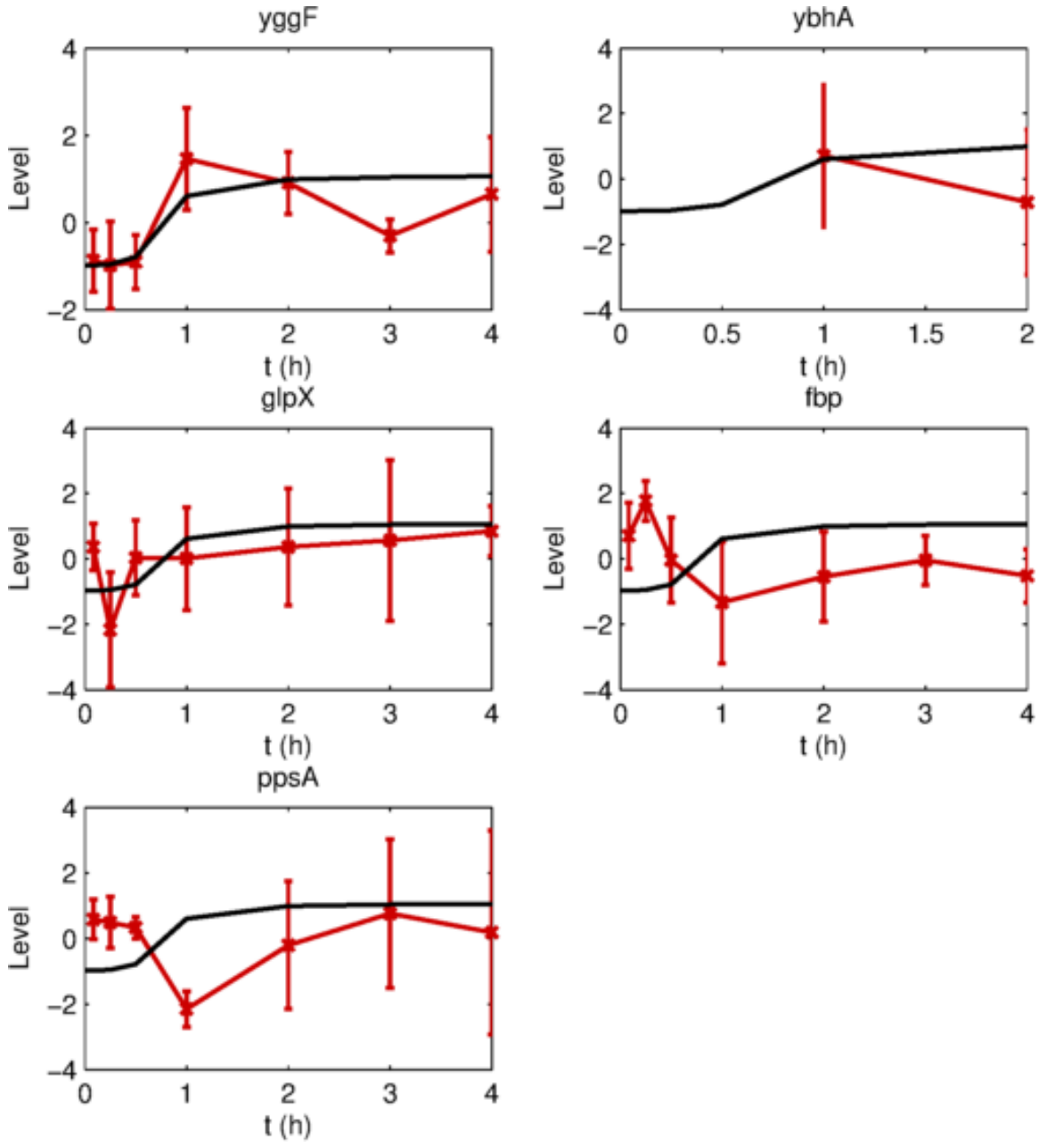

Figure B.2.: Gluconeogenesis pathway gene transcripts compared to acetate u_i variable.

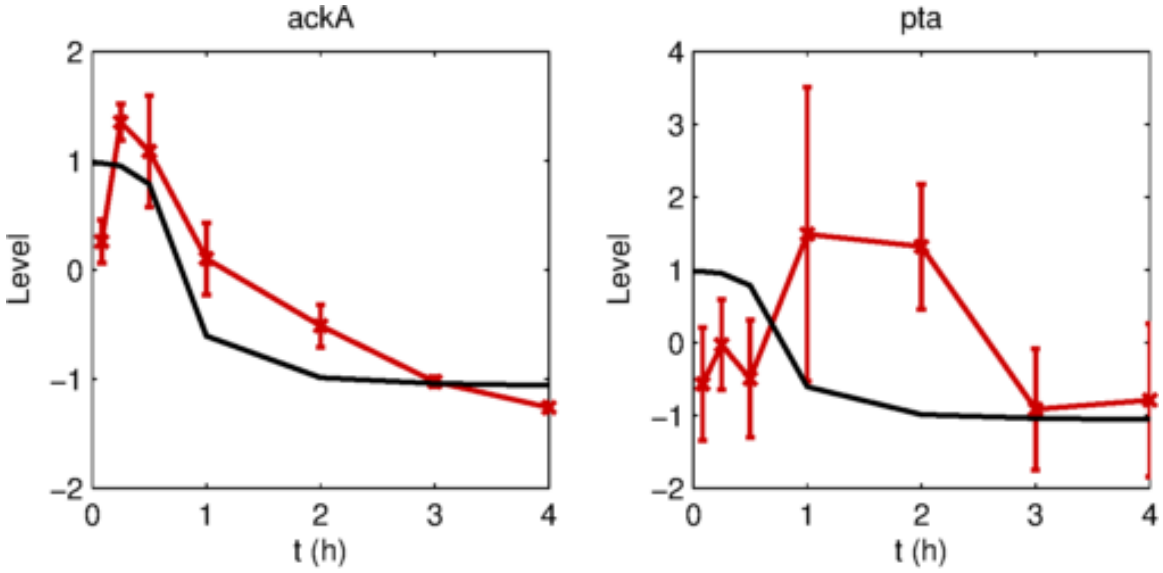

Figure B.3.: Acetate secretion pathway gene transcripts compared to glucose u_i variable.

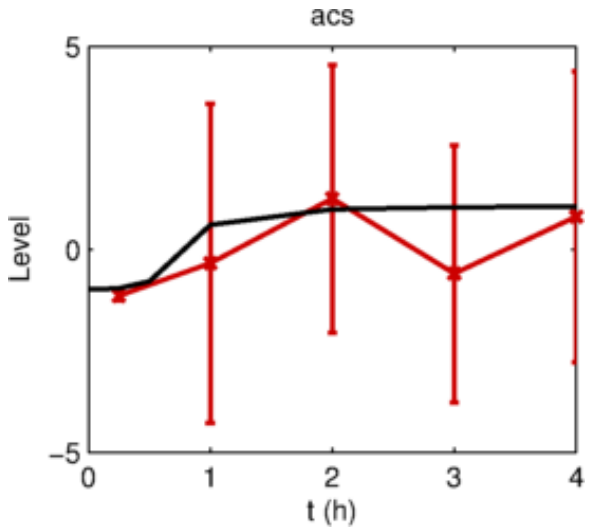

Figure B.4.: Acetate uptake pathway gene transcripts compared to acetate u_i variable.

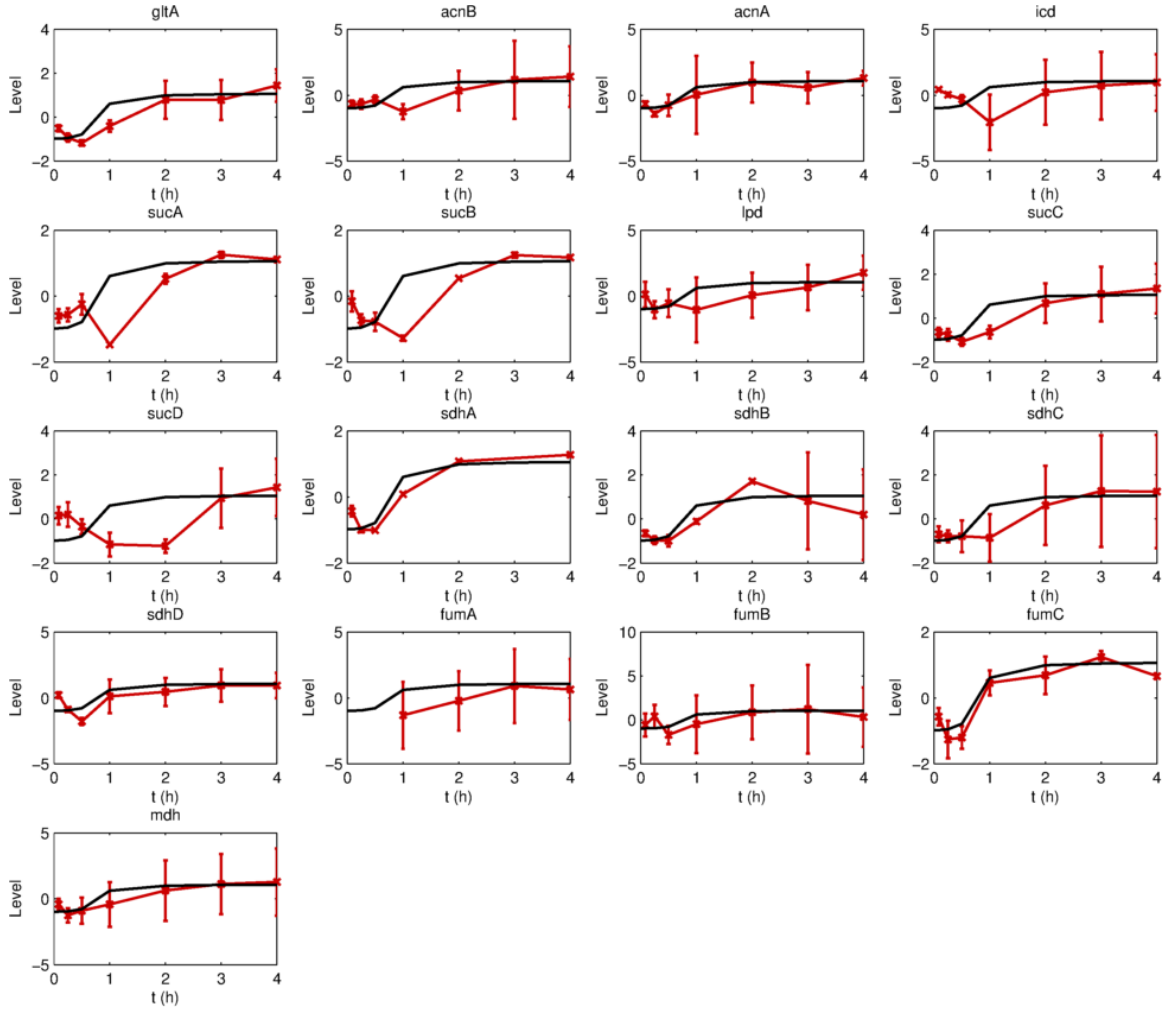

Figure B.5.: TCA cycle gene transcripts compared to acetate u_i variable.

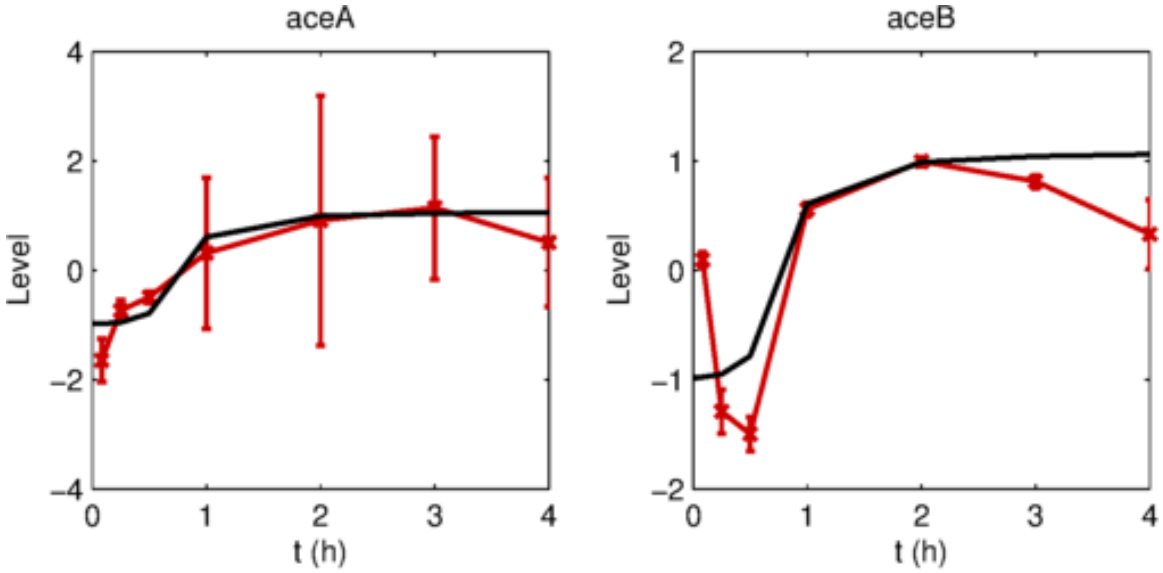

Figure B.6.: Glyoxylate pathway gene transcripts compared to acetate u_i variable.

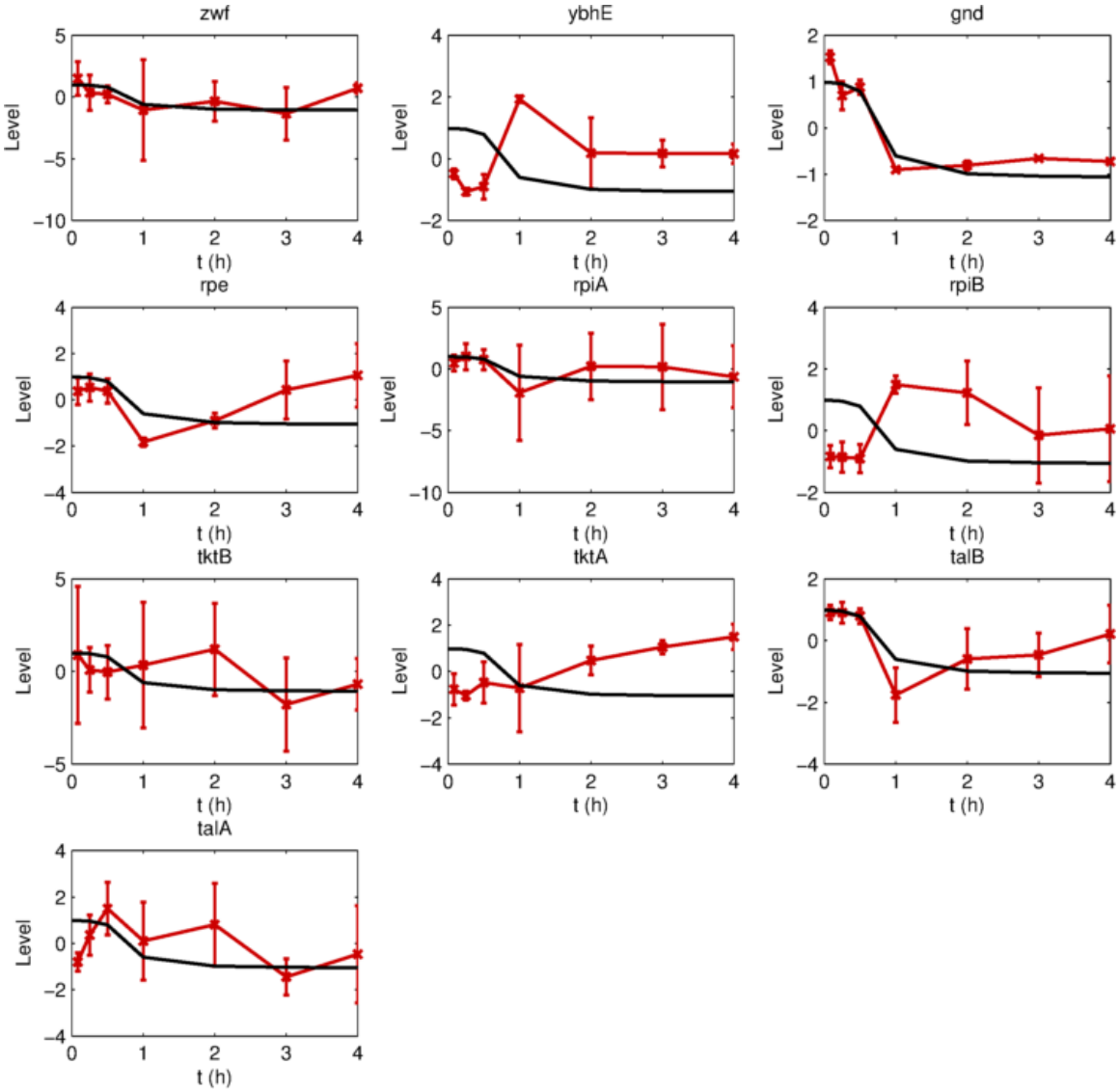

Figure B.7.: Pentose phosphate pathway gene transcripts compared to glucose u_i variable.

C. LISTS OF GENES USED FOR PATHWAY ENRICHMENT ANALYSIS

Table C.1.: Goal Signal Gene List - Glucose-Acetate System

genes						
aceB	aceE	acnA	ada	agaR	agaV	agaW
apaH	araH_2	argG	argH	aroL	arp	artJ
ascB	asd	aspA	aspS	asr	atoB	atoC
atoD	b0298	b0302	b0499	b0669	b0845	b0965
b1173	b1345	b1362	b1443	b1550	b1578	b1604
b1605	b1624	b1640	b1667	b1689	b1691	b1696
b1728	b1729	b1953	b1966	b2326	b2345	b2372
b2383	b2490	b2512	b2817	b2832	b2889	b2896
b2972	b3022	basR	bioH	bisC	btuR	cadC
ccmH	creC	cutC	cysB	cysC	cysE	dbpA
dicB	dsbB	eutG	fabG	fimF	fis	fldA
flgK	flgL	fliL	fliM	fliN	fliZ	fmt
ftsA	ftsK	fucO	fumC	glnD	glnE	glnG
glpB	gltA	gmk	grxA	gsp	hcaA1	hlpA
hofF	hofG	hsdS	htrA	hycE	hycF	hycG
hyfA	hyfE	hypC	idnR	insA_5	insA_6	insB_1
insB_3	ispB	kefC	lgt	lpdA	lpxA	lyxK
malF	mcrA	mdh	mdlA	mdlB	mdoB	menE
menF	menG	mesJ	mhpE	mhpF	mhpR	minC

continued on next page

Table C.1.: *continued*

genes						
moaA	mobB	motB	mraY	mrcB	msrA	murD
mviN	nadA	nadB	nagA	ndh	nrdB	nrdD
nrdG	nrfE	nrfF	nth	nuoB	nuoE	nuoH
nuoL	nupC	nusA	oppC	oppD	osmC	panB
pdxA	pepD	pgk	pgsA	phnE	phnK	phnO
phnP	phrB	pin	plsC	plsX	pmbA	pnp
potA	potF	pphA	ppiA	ppiB	priA	priC
prlA	proA	proV	prpB	prpE	pspD	pssR
ptsI	purD	purL	purN	purT	pyrC	recB
recR	relA	rep	rfaB	rhlB	rhlE	rho
rhsA	rpiR	rplK	rplQ	rseC	sapC	sapD
sapF	sdaA	sdaC	sdhA	sdhD	sdiA	secE
secG	sfa	sfmD	slyD	smg	smpB	sodB
soxR	sppA	sprT	sseA	sucB	tauB	tdcB
tehA	tehB	thiE	thrS	thyA	tnaB	tnaL
tra5_3	trpS	trs5_1	trs5_5	trxC	ttdA	tyrA
tyrB	tyrS	ubiX	uhpT	usg	ushA	uup
uvrA	uvrB	uvrY	uxaA	vacB	wbbK	wcaA
wcaE	wecB	wecD	xthA	xylF	xylG	yabQ
yadE	yadF	yadL	yadQ	yaeE	yafC	yafE
yafN	yafS	yafV	yagF	yagH	yagM	yagP
yagQ	yahB	yahH	yaiD	yaiT	yajG	ybaA
ybaC	ybaD	ybaE	ybaJ	ybaK	ybbS	ybbT
ybbU	ybbV	ybbW	ybbY	ybcM	ybcN	ybcO
ybdE	ybdF	ybdH	ybdS	ybfC	ybfD	ybfE

continued on next page

Table C.1.: *continued*

genes						
ybfF	ybfG	ybfP	ybgC	ybgF	ybgJ	ybgK
ybgL	ybgO	ybgP	ybhD	ybhH	ybhJ	ybhK
ybhL	ybhM	ybhR	ybiA	ybiB	ybiC	ybiF
ybiH	ybiI	ybiM	ybiN	ybiO	ybiU	ybiW
ybiX	ybiY	ycbC	ycbE	ycbG	ycbS	ycbW
ycbY	yccA	ycdC	ycdY	yceE	ycfJ	ycfK
ycfO	ycfP	ycfQ	ycfR	ycfW	ycgE	ychE
yciC	yciI	yciO	yciV	ycjV	ycjW	ydaJ
ydaK	ydaW	ydaY	ydbA_1	ydbP	ydcN	ydeA
ydeD	ydgR	ydiR	yebF	yebU	yedE	yedF
yehD	yejK	yfhG	yfhJ	yfiB	yfiF	yfjA
yfjR	ygcP	ygdD	ygfT	ygfU	yggD	yggE
yggP	yggX	ygiL	ygiM	ygiN	ygiP	ygjO
yhaH	yhbJ	yhbQ	yhbW	yhbX	yhdA	yhdE
yhdX	yhdY	yheR	yhfM	yhfR	yhhA	yhjM
yhjU	yi22_5	yi81_2	yi91b	yiaL	yiaN	yicG
yicK	yidP	yidW	yigG	yigK	yigR	yihT
yiiE	yiiG	yjbF	yjdL	yjeT	yjfJ	yjgF
yjgR	yjhH	yjiD	yjiZ	yjjA	yjjB	yjjW
ylaC	ymbA	yneH	ynhC	yohL	ypjE	yqeJ
yqjF	yraH	yraJ	yraL	yrbC	yrbD	yrbG
yrbK	yrfF	ytfF	zwf			

Table C.2.: Glucose-Lactose System Goal Signal Gene List

murG	murF	murD
araA	leuC	hepA
carB	yadF	ilvI
aceE	ampD	ileS
araB	apaH	ftsA
ftsI	leuA	ampE
hofB	yaaH	yadG
mraY	thrA	caiT
leuD	yaaI	yabO
yabP	htgA	ftsZ
yacF	speD	guaC
yi82_1	yacK	yabM
murC	yacG	
yabK	talB	
b0100	hofC	

Table C.3.: Prostaglandin Goal Signal Gene List (Entrez IDs / p-value < 0.05)

11542	11861	12371	12421	12686
12768	12771	13032	13132	14026
14130	14131	14313	14544	15368
15468	16181	16190	16409	16416
16438	16633	16911	17000	17112
17118	17119	17392	17480	17768
19079	19164	20293	20405	20479
20619	20907	21816	22142	22158
26941	26951	27362	28018	52004
52502	52662	52808	53599	53622
55963	56045	56177	56389	56737
58248	64382	66146	66251	66607
66898	66943	67098	67510	67588
67946	69065	69076	69608	69710
69743	69772	69959	70829	72245
72925	73710	74018	74150	74198
74603	74781	75871	76457	77125
77328	79555	80289	83602	83679
101497	105278	106200	106326	106572
107321	107934	108112	108954	109815
114601	192650	192652	208228	211496
212919	217664	226245	226409	230088
230784	232370	234199	241296	245403
258280	258940	269346	319370	327957
380686	380712	381484	382113	634881

Table C.4.: Prostaglandin Goal Signal Gene List (Entrez IDs / p-value < 0.10)

genes								
11542	11749	11812	11861	11977	12370	12371	12421	12481
12686	12768	12771	12928	12953	12986	12988	13026	13032
13132	13390	13730	13805	13830	14026	14038	14070	14130
14131	14254	14313	14544	14977	15368	15421	15468	16181
16190	16328	16409	16416	16438	16443	16633	16859	16869
16911	16997	17000	17087	17112	17118	17119	17392	17480
17533	17768	17979	18550	18647	18704	18715	18788	19012
19079	19082	19108	19164	19219	19766	20229	20293	20353
20405	20439	20479	20555	20619	20773	20907	21816	21887
22142	22158	22324	23796	24056	26941	26951	27081	27261
27362	27381	28018	28040	28146	50523	52004	52055	52120
52502	52662	52808	53378	53599	53622	53791	54170	54713
55963	56045	56177	56380	56389	56484	56632	56722	56737
57435	57746	57781	57912	58203	58248	58998	59007	60361
64294	64380	64382	65973	66146	66251	66503	66607	66684
66816	66898	66943	66989	66994	66998	67098	67171	67198
67231	67457	67510	67588	67608	67946	67976	68038	68118
68205	68774	68952	69065	69076	69608	69710	69718	69743
69772	69834	69959	70231	70829	71782	71967	72245	72344
72925	73393	73710	73808	73827	74018	74150	74198	74568
74603	74781	74868	75018	75288	75871	76281	76457	76775
77125	77328	77573	78412	78610	78611	78830	79555	80289
80795	83397	83602	83679	83704	83984	93692	93888	94220
98878	100226	101489	101497	103806	104110	105278	106200	106326

continued on next page

Table C.4.: *continued*

genes								
106572	107321	107373	107934	108112	108954	109019	109674	109815
110920	114601	114716	170644	170744	192650	192652	207818	208228
208647	210044	210711	211329	211496	212919	213438	214855	216161
217664	219144	224432	226245	226409	227738	228775	229600	229707
230088	230784	231946	232333	232370	233571	233651	234199	236312
241296	241624	242737	244416	245403	246256	258280	258381	258727
258940	259003	263876	268752	269132	269346	270160	319355	319370
319622	319922	327957	329977	353190	380686	380712	380921	381356
381484	381724	382113	382639	433809	544905	545474	634881	654440

VITA

VITA

Frank DeVilbiss was born in Altoona, PA in 1989 and spent most of his childhood in Carmel, IN. After graduating from Carmel High School in 2007, he went to Purdue University for his B.S. in chemical engineering which he received in May 2011. He then continued at Purdue in the School of Chemical Engineering for his PhD.